Analysis and Topology

University of Cambridge Part IB Mathematical Tripos

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Contents

1 Uniform Convergence

1.1 Uniform Convergence

Definition 1.1. Let E be a set and $f_n : E \to \mathbb{R}$ be a sequence of functions. The sequence is said to *converge pointwise* to another function $f : E \to \mathbb{R}$ if

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for all $x \in E$.

This is an easy definition that is simple to check. However, there is a problem. Ideally, we want to deduce properties of f from properties of f_n. For example, it would be great if the continuity of all f_n implies continuity of f, and similarly for integrability and values of derivatives and integrals. However, it turns out we cannot. The notion of pointwise convergence is too weak.

Example. Consider a sequence of functions $f_n : [-1, 1] \to \mathbb{R}$ defined by $f_n = x^{1/(2n+1)}$. These are all continuous, but their pointwise limit function is

$$f_n(x) \to f(x) = \begin{cases} 1 & x \in (0, 1] \\ 0 & x = 0 \\ -1 & x \in [-1, 0) \end{cases}$$

which is not continuous. The continuity of functions is not preserved.

Example. Consider a sequence of functions $f_n : [0, 1] \to \mathbb{R}$ be the piecewise linear function formed by joining the points $(0, 0)$, $(\frac{1}{n}, n)$, $(\frac{2}{n}, 0)$, $(1, 0)$. The pointwise limit of this function is $f_n(x) \to f(x) = 0$. However, we have

$$\int_0^1 f_n(x)\,\mathrm{d}x = 1 \text{ for all } n, \text{ but } \int_0^1 f(x)\,\mathrm{d}x = 0\,.$$

The limit of the integral is not the integral of the limit.

Example. Let $f_n : [0, 1] \to \mathbb{R}$ be

$$f_n(x) = \begin{cases} 1 & n!x \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases},$$

which all have finitely many discontinuities, so are Riemann integrable. However, its limit is

$$f_n(x) \to f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases},$$

which is not integrable. So the integrability of a function is not preserved by pointwise limits.

We need a stronger notion of convergence without being too trivial. In the above example, we may see that the problem of pointwise convergence is that different points converge at different rates to f – it can be arbitrarily slow for some points. We need f_n to converge to f at the same pace.

Definition 1.2. A sequence of functions $f_n : E \to \mathbb{R}$ is said to *converge uniformly* to $f : E \to \mathbb{R}$ if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$, $|f_n(x) - f(x)| < \epsilon$ for all $x \in E$.

Remark. Uniform convergence naturally implies pointwise convergence.

The definition does not require E to be a subset of $\mathbb{R}$, but many of our theorems will require it to be, or else we cannot sensibly integrate or differentiate our functions.

We will now see that the uniform limit function retains certain properties from the original sequence.

Theorem 1.3 (Continuity of the uniform limit). Let $\Omega \subseteq \mathbb{R}$. Given a sequence of continuous functions $f_n : \Omega \to \mathbb{R}$, if $f_n \to f$ uniformly on Ω, then f is continuous.

Proof. We will show that f is continuous at an arbitrary $a \in \Omega$. Given $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ and $x \in \Omega$ we have $|f_n(x) - f(x)| < \epsilon/3$. Then since f_n is continuous, there exists $\delta > 0$ such that $|x - a| < \delta$ implies $|f_n(x) - f_n(a)| < \epsilon/3$. Then if $x \in \Omega$ and $|x - a| < \delta$, we have

$$|f(x) - f(a)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| \leq \epsilon$$

as required. □

Remark. Another way of thinking this is the uniform convergence allows us to swap the limit:

$$\lim_{x\to a} f(x) = \lim_{x\to a} \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \lim_{x\to a} f_n(x) = f(a)\,,$$

which implies continuity at $x = a$.

Lemma 1.4 (Boundedness of the uniform limit). Let $f_n \to f$ uniformly on a set Ω. If f_n is bounded for every n, then so is f.

Proof. Fix some $n \in \mathbb{N}$ such that for all $x \in \Omega$, we have $|f_n(x) - f(x)| < 1$. Then since f_n is bounded, there is an $M \in \mathbb{R}$ such that $|f_n(x)| \leq M$ for all $x \in \Omega$. Thus

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq M + 1$$

so f is bounded. □

Theorem 1.5 (Integral of the uniform limit). Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of integrable functions. If $f_n \to f$ uniformly on $[a, b]$ then f is integrable and

$$\int_a^b f_n \,\mathrm{d}x \to \int_a^b f \,\mathrm{d}x\ .$$

Proof. Since every f_n is integrable, they must be all bounded and hence f is also bounded. It suffices to check that Riemann's criterion is satisfied.

Given $\epsilon > 0$, we have $N \in \mathbb{N}$ such that for all $n \geq N$, $x \in [a, b]$ implies $|f_n(x) - f(x)| < \epsilon$. Then since f_n is integrable, there exists a dissection $\mathcal{D} = \{x_0, x_1, \ldots, x_m\}$ of $[a, b]$ such that $U_\mathcal{D}(f_n) - L_\mathcal{D}(f_n) < \epsilon$.

Now for each $k \in \{1, \ldots, m\}$ and any $x, y \in [x_{k-1}, x_k]$, we have

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq |f_n(x) - f_n(y)| + 2\epsilon\,. \end{aligned}$$

Therefore, the difference between the supremum and infimum of f in $[x_{k-1}, x_k]$ can be identified as

$$\sup_{x,y\in[x_{k-1},x_k]} |f(x) - f(y)| \leq \sup_{x,y\in[x_{k-1},x_k]} |f_n(x) - f_n(y)| + 2\epsilon\,.$$

We can multiply by $(x_k - x_{k-1})$ and sum over k to get

$$U_\mathcal{D}(f) - L_\mathcal{D}(f) \leq U_\mathcal{D}(f_n) - L_\mathcal{D}(f_n) + 2\epsilon(b - a) \leq \epsilon(2(b - a) + 1)\,,$$

so f is integrable.

The integral can be easily identified,

$$\left| \int_a^b f_n \,\mathrm{d}x - \int_a^b f \,\mathrm{d}x \right| \leq \int_a^b |f_n - f| \,\mathrm{d}x \leq (b - a) \sup_{s\in[a,b]} |f_n - f| \to 0\,.$$

□

Remark. Similarly, this can be seen as changing the place of the limit:

$$\int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \lim_{n\to\infty}\int_a^b f_n(x)\,\mathrm{d}x\ .$$

Theorem 1.6 (Term-wise integration). Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of integrable functions. Then if $\sum_{n=1}^\infty f_n(x)$ converges uniformly on $[a,b]$, the function $x \mapsto \sum_{n=1}^\infty f_n(x)$ is integrable and

$$\int_a^b \sum_{n=1}^\infty f_n(x)\,\mathrm{d}x = \sum_{n=1}^\infty \int_a^b f_n(x)\,\mathrm{d}x\ .$$

Proof. Let g_n be the sequence of partial sums. g_n converges uniformly so we can apply the previous theorem. □

However, the relationship between uniform convergence and differentiability is more subtle. The uniform limit of differentiable functions need not be differentiable. Even if it were, the limit of the derivative is not necessarily the same as the derivative of the limit, even if we just want pointwise convergence of the derivative.

Example. Let $f_n, f : [-1,1] \to \mathbb{R}$ be

$$f_n(x) = |x|^{1+1/n}\,,\ f(x) = |x|\,.$$

Then $f_n \to f$ uniformly. Each f_n is differentiable throughout $[-1,1]$ - this is obvious at $x \neq 0$. At $x = 0$, the derivative is

$$f_n'(0) = \lim_{x\to 0}\frac{f_n(x) - f_n(0)}{x} = \lim_{x\to 0}\operatorname{sgn}(x)|x|^{1/n} = 0\,.$$

However, the uniform limit f is not differentiable at $x = 0$.

Example. Let

$$f_n(x) = \frac{\sin nx}{\sqrt{n}}$$

for $x \in \mathbb{R}$. Then

$$\sup_{x\in\mathbb{R}}|f_n(x)| \le \frac{1}{\sqrt{n}} \to 0\,,$$

so $f_n(x) \to f = 0$ uniformly on $\mathbb{R}$. However, the derivative is

$$f_n'(x) = \sqrt{n}\cos nx\,,$$

which does not converge to $f' = 0$.

We need a condition even stronger than uniform convergence to preserve differentiability and derivative.

Theorem 1.7 (Term-wise differentiation). Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of continuously differentiable functions. If $\exists c \in [a,b]$ such that $\sum_{i=1}^\infty f_i(c)$ converges and $\sum_{i=1}^\infty f_i'(x)$ converges uniformly on $[a,b]$, then $\sum_{i=1}^\infty f_i$ also converges uniformly to a continuously differentiable function f, and

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{j=1}^\infty f_j(x)\right) = \sum_{j=1}^\infty f_j'(x)\,.$$

Proof. Let $g(x) = \sum_{k=1}^\infty f_k'(x)$ for $x \in [a,b]$. The general idea is that we want to solve the differential equation $f' = g$ subjected to the initial condition $f(c) = \sum_{n=1}^\infty f_n(c)$. Let $\lambda = \sum_{n=1}^\infty f_n(c)$ and define $f : [a,b] \to \mathbb{R}$ by

$$f(x) = \lambda + \int_c^x g(t)\,\mathrm{d}t\ .$$

Note that g is integrable: $\sum_{k=1}^{\infty} f_k' \to g$ uniformly implies that g is continuous and hence integrable. By the fundamental theorem of calculus, $f' = g$ and $f(c) = \lambda$. So we have found such an f that satisfies the conditions set out. All that remains is to prove the uniform convergence of $\sum_{k=1}^{\infty} f_k \to f$.

Also by the fundamental theorem of calculus,

$$f_k(x) = f_k(c) + \int_c^x f_k'(t)\,\mathrm{d}t\ .$$

For $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|\lambda - \sum_{k=1}^{n} f_k(c)| < \epsilon$ and $|g(t) - \sum_{k=1}^{n} f_k'(t)| < \epsilon$ for all $n \geq N$. Then we have

$$\begin{aligned}\left|f(x) - \sum_{k=1}^{n} f_k(x)\right| &= \left|\lambda + \int_c^x g(t)\,\mathrm{d}t - \sum_{k=1}^{n}\left(f_k(c) + \int_c^x f_k'(t)\,\mathrm{d}t\right)\right| \\ &\leq \left|\lambda - \sum_{k=1}^{n} f_k(c)\right| + \left|\int_c^x \left(g(t) - \sum_{k=1}^{n} f_k'(t)\right)\mathrm{d}t\right| \\ &\leq \epsilon + |x-c|\epsilon \\ &\leq \epsilon(b-a+1)\,.\end{aligned}$$

□

1.2 The General Principle of Convergence

Definition 1.8. A sequence of functions f_n on a set Ω is *uniformly Cauchy* if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall m, n > N$ and $x \in \Omega$, we have

$$|f_m(x) - f_n(x)| < \epsilon\,.$$

Theorem 1.9 (General principle of uniform convergence). Let $f_n : \Omega \to \mathbb{R}$ be a sequence of functions. It converges uniformly to some $f : \Omega \to \mathbb{R}$ if and only if it is uniformly Cauchy.

Proof.

($\Rightarrow$) Let $f_n \to f$ uniformly. Then given $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$,

$$\sup_{x\in\Omega} |f_n(x) - f(x)| < \frac{\epsilon}{2}\,.$$

Then if $n, m > N$, for all $x \in \Omega$, we have

$$|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f_m(x) - f(x)| < \epsilon\,,$$

so f_n is uniformly Cauchy.

($\Leftarrow$) Now let f_n be uniformly Cauchy, then $\forall x \in \Omega$, the sequence of real numbers $f_n(x)$ is Cauchy and hence convergent. Let

$$f(x) = \lim_{n\to\infty} f_n(x)\,.$$

We want to show that $f_n \to f$ uniformly. Given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that $\forall n, m > N$ and $\forall x \in \Omega$, we have $|f_n(x) - f_m(x)| < \epsilon/2$. Fix some $x \in \Omega$, then choose m such that $|f_m(x) - f(x)| < \epsilon/2$. Then

$$|f_n(x) - f(x)| \leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \epsilon\,.$$

Note that this result is independent of x, although the choice of m does depend on x. Hence $f_n \to f$ uniformly. □

From this, we have our important Weierstrass M-test for uniform convergence.

Theorem 1.10 (Weierstrass M-test). Let $f_n : \Omega \to \mathbb{R}$ be a sequence of functions. Assume that $\forall n \in \mathbb{N}$, $\exists M_n \in \mathbb{R}^+$ such that $\sup_{x\in\Omega} |f_n(x)| \leq M_n$. If $\sum_{n=1}^{\infty} M_n$ converges, then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on Ω.

Proof. Let $F_n(x) = \sum_{k=1}^{n} f_k(x)$. For $x \in \Omega$ and $n \geq m$, we have

$$|F_n(x) - F_m(x)| \leq \sum_{k=m+1}^{n} |f_k(x)| \leq \sum_{k=m+1}^{n} M_k \,.$$

Now given $\epsilon > 0$, we can choose $N \in \mathbb{N}$ such that $\sum_{k=N+1}^{\infty} M_k < \epsilon$. Then for every $x \in \Omega$ and $n \geq m \geq N$, we have

$$|F_n(x) - F_m(x)| \leq \sum_{k=m+1}^{\infty} M_k < \epsilon \,.$$

F_n is uniformly Cauchy and hence uniformly convergent. □

1.3 Power Series

Definition 1.11. A complex open disk $D(a, R)$ is the set

$$D(a, R) = \{z \in \mathbb{C} \mid |z - a| < R\} \,,$$

where $a \in \mathbb{C}$ and $R \in \mathbb{R}^+$.

We are interested in whether a complex power series converge uniformly within its radius of convergence.

Example. Consider the power series

$$\sum_{n=1}^{\infty} \frac{z^n}{n^2} \,,$$

which has a radius of convergence $R = 1$. Let $z \in D(0, 1)$. Then $|f_n(z)| \leq 1/n^2$. We have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

is convergent, so by the Weierstrass M-test, this power series converges uniformly on $D(0, 1)$.

Example. Consider $\sum_{n=0}^{\infty} z^n$ which has the radius of convergence $R = 1$. It has bounded partial sums for $z \in D(0, 1)$, but it converges to $1/(1 - z)$, which is unbounded on $D(0, 1)$. Hence, the convergence is not uniform, otherwise the boundedness will be preserved.

We can see that the uniform convergence fails right at the boundary. We can keep the uniform convergence if we move slightly away.

Theorem 1.12 (Uniform convergence of power series). Let $\sum_{n=0}^{\infty} c_n(z - a)^n$ be a complex power series with a radius of convergence R. Then $\forall r \in (0, R)$, the power series converges uniformly on $D(a, r)$.

Proof. Fix some $w \in \mathbb{C}$ such that $|a - w| \in (r, R)$. Define $\rho = \frac{r}{|w-a|}$ so that $\rho \in (0, 1)$. Since $\sum_{n=0}^{\infty} c_n(w - a)^n$ converges, we have $c_n(w - a)^n \to 0$ as $n \to \infty$. Thus there exists some $M \in \mathbb{R}^+$ such that $|c_n(w - a)^n| \leq M$ for all $n \in \mathbb{N}$.

Now if we take $z \in D(a,r)$ and $n \in \mathbb{N}$, we have

$$\begin{aligned} |c_n(z-a)^n| &= |c_n(w-a)^n| \cdot \left(\frac{|z-a|}{|w-a|}\right)^n \\ &\leq M\left(\frac{r}{|w-a|}\right)^n \leq M\rho^n \,. \end{aligned}$$

Then since $\sum_{n=0}^{\infty} M\rho^n$ is convergent, by the Weierstrass M-test, we have that the power series is converging uniformly on $z \in D(a,r)$. □

Corollary. Let $\sum_{n=0}^{\infty} c_n(z-a)^n$ be a complex power series with radius of convergence R. Then

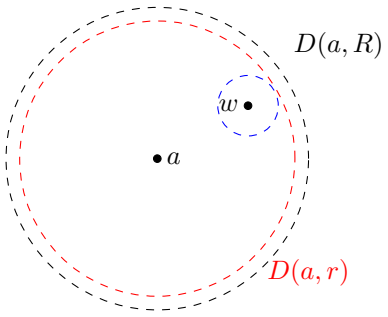
(i) the power series is continuous within its radius of convergence;

(ii) the derivative of the power series is

$$\sum_{n=1}^{\infty} nc_n(z-a)^{n-1} \,.$$

It is not guaranteed that a power series converge uniformly on its whole disk of convergence. However, we have shown that we can get arbitrarily close to its radius of convergence, and we can still have some amount of uniform convergence.

Consider a power series $\sum_{n=0}^{\infty} c_n(z-a)^n$ with radius of convergence R and a point $w \in D(a,R)$. We can always find some r such that $|w-a| < r < R$, so $\exists \delta > 0$ such that $|w-a| + \delta < r$. Therefore, we can always find an open disk around w with radius δ such that it is completely contained within $D(a,r)$.



We then have $D(w,\delta) \subset D(a,r)$, where the power series is guaranteed to converge uniformly.

This inspires a helpful definition.

Definition 1.13. A subset U of $\mathbb{C}$ is *open* if for all $w \in \mathbb{U}$, there exists $\delta > 0$ such that $D(w,\delta) \subset U$.

From our discussion above, it is clear that an open disk is open.

Definition 1.14. Let $U \subset \mathbb{C}$ be open and f_n be a sequence of functions on U. We say that f_n *converges locally uniformly* on U if for all $w \in U$ there exists some $\delta > 0$ such that f_n converges uniformly on $D(w,\delta) \subset U$.

We can then formalise our discussion above into the following theorem.

Theorem 1.15 (Local uniform convergence of power series). A complex power series centred at a with radius of convergence R converges locally uniformly on $D(a,R)$.

2 Uniform Continuity and Integrability

2.1 Uniform Continuity

Recall the standard notion of continuity.

Definition 2.1. Let $\Omega \subseteq \mathbb{C}$ and $f : \Omega \to \mathbb{C}$ (or $\mathbb{R}$). We say that f is *continuous* at $a \in \Omega$ if given any $\epsilon > 0$ we can find a $\delta > 0$ such that $|f(a) - f(x)| < \epsilon$ for all $x \in \Omega, |x - a| < \delta$.

We say f is *continuous* if it is continuous at every $a \in \Omega$.

In our above definition, we allow δ to depend on both ϵ and a. Similar to how we define uniform convergence, we can analogously define uniform continuity.

Definition 2.2. Let $\Omega \subset \mathbb{C}$ and $f : \Omega \to \mathbb{C}$ (or $\mathbb{R}$). We say that f is *uniformly continuous* if given any $\epsilon > 0$, we can find $\delta > 0$ such that $\forall x, y \in \Omega$, $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$.

It is easy to see that uniform continuity implies continuity, but the converse does not hold in general.

Theorem 2.3 (Heine–Cantor theorem). Let $f : [a, b] \to \mathbb{C}$ be continuous. Then f is uniformly continuous.

Proof. Suppose such f is not uniformly continuous. Then $\exists \epsilon > 0$ such that $\forall \delta > 0$, there is some x, y with $|x - y| < \delta$ such that $|f(x) - f(y)| \geq \epsilon$.

Take $\delta = 1/n$. We can find some sequences $x_n, y_n \in [a, b]$ such that $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \epsilon$. Then by Bolzano–Weierstrass theorem, we can find some convergent subsequence $x_{n_j} \to x$ for some $x \in [a, b]$. But then $\left|x_{n_j} - y_{n_j}\right| < 1/n_j$ for all j, so we must have $y_{n_j} \to x$ as well.

Then $\left|f(x_{n_j}) - f(y_{n_j})\right| \geq \epsilon$ for every j, which implies that $f(x_{n_j})$ and $f(y_{n_j})$ cannot converge to the same point. But by continuity of f, we must have $f(x_{n_j}) \to f(x)$ and $f(y_{n_j}) \to f(x)$, which is a contradiction. Thus f must be uniformly continuous. □

2.2 Riemann Integrability

Theorem 2.4 (Integrability and uniform continuity). Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is Riemann integrable.

Proof. Given $\epsilon > 0$, since f is uniformly continuous, there is some $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/(b - a)$ whenever $|x - y| < \delta$ and $x, y \in [a, b]$.

Now choose some integer $n \geq (b - a)/\delta$, and define the dissection $\mathcal{D} = \{x_0, \ldots, x_n\}$ with $x_i = a + i(b - a)/n$. Then we have

$$\sup_{x \in [x_{j-1}, x_j]} f(x) - \inf_{x \in [x_{j-1}, x_j]} f(x) \leq \frac{\epsilon}{(b - a)}$$

for all $1 \leq j \leq n$, and thus

$$\begin{aligned} U_{\mathcal{D}}(f) - L_{\mathcal{D}}(f) &= \sum_{j=1}^{n} (x_j - x_{j-1}) \left[\sup_{x \in [x_{j-1}, x_j]} f(x) - \inf_{x \in [x_{j-1}, x_j]} f(x) \right] \\ &\leq \sum_{j=1}^{n} \frac{b - a}{n} \frac{\epsilon}{b - a} = \epsilon \,, \end{aligned}$$

and thus f is Riemann integrable by Riemann criterion for integrability. □

*Non-examinable fun

Below are some remarkable results that are non-examinable.

Theorem 2.5 (Weierstrass approximation theorem). If $f : [0,1] \to \mathbb{R}$ is continuous, then there exists a sequence of polynomials p_n such that $p_n \to f$ uniformly. The sequence can be given by the *Bernstein polynomial*:

$$p_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} .$$

Remark. There are many other ways to construct a sequence of polynomials converging uniformly to a continuous function f.

Proof. Denote

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k} .$$

We can derive a few facts about these functions. Clearly, $p_{n,k}(x) \geq 0 \ \forall x \in [0,1]$. Also, by the binomial theorem

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n ,$$

so we get

$$\sum_{k=0}^{n} p_{n,k}(x) = 1 .$$

Differentiating the binomial theorem with respect to x and putting $y = 1 - x$ gives

$$\sum_{k=0}^{n} \binom{n}{k} k x^{k-1} (1-x)^{n-k} = n .$$

Multiply by x gives

$$\sum_{k=0}^{n} \binom{n}{k} k x^k (1-x)^{n-k} = nx ,$$

or equivalently

$$\sum_{k=0}^{n} k p_{n,k}(x) = nx .$$

Differentiating the binomial theorem twice gives

$$\sum_{k=0}^{n} k(k-1) p_{n,k}(x) = n(n-1)x^2 .$$

Adding these two results gives

$$\sum_{k=0}^{n} k^2 p_{n,k}(x) = n^2 x^2 + nx(1-x) .$$

We will combine our results in a weird way:

$$\begin{aligned} \sum_{k=0}^{n} (nx-k)^2 p_{n,k}(x) &= n^2x^2 - 2nx \cdot nx + n^2x^2 + nx(1-x) \\ &= nx(1-x) . \end{aligned}$$

Now given ϵ, since f is continuous on $[0,1]$, it is uniformly continuous. So pick a δ such that $|f(x)-f(y)|<\epsilon$ whenever $|x-y|<\delta$. Then for each fixed x, we can write

$$\begin{aligned}
|p_n(x)-f(x)| &= \left|\sum_{k=0}^{n}\left(f\left(\frac{k}{n}\right)-f(x)\right)p_{n,k}(x)\right| \\
&\leq \sum_{k=0}^{n}\left|f\left(\frac{k}{n}\right)-f(x)\right|p_{n,k}(x) \\
&= \sum_{k:\left|x-\frac{k}{n}\right|<\delta}\left|f\left(\frac{k}{n}\right)-f(x)\right|p_{n,k}(x)+\sum_{k:\left|x-\frac{k}{n}\right|\geq\delta}\left|f\left(\frac{k}{n}\right)-f(x)\right|p_{n,k}(x) \\
&\leq \epsilon\sum_{k=0}^{n}p_{n,k}(x)+2\sup_{x\in[0,1]}|f(x)|\sum_{k:\left|x-\frac{k}{n}\right|\geq\delta}p_{n,k}(x) \\
&= \epsilon+2\sup_{x\in[0,1]}|f(x)|\frac{1}{\delta^2}\sum_{k:\left|x-\frac{k}{n}\right|\geq\delta}\left(x-\frac{k}{n}\right)^2p_{n,k}(x) \\
&\leq \epsilon+2\sup_{x\in[0,1]}|f(x)|\frac{1}{\delta^2}\sum_{k=0}^{n}\left(x-\frac{k}{n}\right)^2p_{n,k}(x) \\
&= \epsilon+\frac{2\sup|f|}{\delta^2n^2}nx(1-x) \\
&\leq \epsilon+\frac{\sup|f|}{2\delta^2n}\,.
\end{aligned}$$

Hence given any ϵ and δ, we can pick n sufficiently large such that $|p_n(x)-f(x)|<2\epsilon$. This is independent of x so the convergence is uniform. □

We may want to explore more on integrability. We know that a function is integrable if it is continuous. But it need not be. It could be discontinuous at finitely many points and still be integrable. If it has countably many discontinuities, then we still can integrate it. How many points of discontinuity can we accommodate if we want to keep integrability?

Definition 2.6. A subset $A\subseteq\mathbb{R}$ is said to have *(Lebesgue) measure zero* if for any $\epsilon>0$, there exists a countable (possibly infinite) collection of open intervals I_j such that

$$A\subseteq\bigcup_{j=1}^{\infty}I_j\,,$$

and

$$\sum_{j=1}^{\infty}|I_j|<\epsilon\,,$$

where $|I_j|$ is the length of the interval.

Examples.

(i) The empty set has measure zero.

(ii) Any finite set has measure zero.

(iii) Any countable set has measure zero. If $A=\{a_0,a_1,\dots\}$, take

$$I_j=\left(a_j-\frac{\epsilon}{2^{j+1}},a_j+\frac{\epsilon}{2^{j+1}}\right).$$

Then A is contained in the union and the sum of lengths is ϵ.

(iv) A countable union of sets of measure zero has measure zero, using a similar proof strategy as above.

(v) Any non-trivial interval does not have measure zero.

(vi) The *Cantor set*, despite being uncountable, has a measure zero. The Cantor set is constructed as follows: start with $C_0 = [0, 1]$. Remove the middle third $(\frac{1}{3}, \frac{2}{3})$ to obtain $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Removing the middle third of each segment again to obtain

$$C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{3}{9}\right] \cup \left[\frac{6}{9}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

Continue iteratively by removing the middle third of each part. Define

$$C = \bigcap_{n=0}^{\infty} C_n \,,$$

which is the Cantor set. Since each C_n consists 2^n disjoint intervals of length $1/3^n$, the total length of the segments of C_n is $(\frac{2}{3})^n \to 0$. So we can cover C by arbitrarily small union of intervals. Hence the Cantor set has measure zero.

Theorem 2.7 (Lebesgue's theorem on the Riemann integral). Let $f : [a, b] \to \mathbb{R}$ be a bounded function, and let $\mathcal{D}_f$ be the set of points of discontinuities of f. Then f is Riemann integrable if and only if $\mathcal{D}_f$ has measure zero.

3 Metric Spaces

3.1 Metric Spaces

In $\mathbb{R}$ and $\mathbb{C}$, we measure the 'closeness' between two points using $|x-y|$. It has an important property: triangle inequality

$$|x-z| \leq |x-y| + |y-z|\,.$$

The triangle inequality acts more or less as saying closeness is a transitive property.

We can abstract away naturally from the absolute value into a more general measure of distance with this property.

Definition 3.1. Let M be a set. A *metric* on M is a function $d : M \times M \to \mathbb{R}_{\geq 0}$ with the following properties:

1. *Positivity.* $\forall x, y \in M$, $d(x,y) \geq 0$, with equality if and only if $x = y$.
2. *Symmetry.* $\forall x, y \in M$, $d(x,y) = d(y,x)$.
3. *Triangle inequality.* $\forall x, y, z \in M$, $d(x,z) \leq d(x,y) + d(y,z)$.

The pair (M,d) is called a *metric space*.

Examples.

(i) Let M be $\mathbb{R}$ or $\mathbb{C}$, then $d(x,y) = |x-y|$ is the usual metric on those sets.

(ii) The *Euclidean metric* is the usual metric in $\mathbb{R}^n$, defined in component form as

$$d(v,w) = |v-w| = \sqrt{\sum_{i=1}^{n} (v_i - w_i)^2}\,.$$

(iii) Let M be any set, then

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

defines the *discrete metric*, and (M,d) is a *discrete metric space*.

(iv) Let G be a group generated by $S \subset G$, where $e \notin S$ and $x \in S \implies x^{-1} \in S$. Then

$$d(x,y) = \min\{n \geq 0 \mid \exists s_1, \ldots, s_n \in S \text{ such that } y = x s_1 \ldots s_n\}$$

defines the *word metric*.

(v) Let p be a prime. Then

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ p^{-n} & \text{otherwise,} \end{cases}$$

where n is the greatest power of p in the prime factorisation of $|x-y|$, defines a metric on $\mathbb{Z}$ known as the *p-adic metric*.

(vi) Let $M = \mathbb{R}^2$, then

$$d(x,y) = \begin{cases} |x-y| & \text{if } x = ky \\ |x| + |y| & \text{otherwise,} \end{cases}$$

defines a metric on M, where $|\cdot|$ is the Euclidean metric on $\mathbb{R}^2$. This is known as the *British railway metric*.

To explain the name of this metric, think of Britain with London as the origin. Since the railway system is far from optimal, all trains go through London. For example, if you want to go from Oxford to Cambridge (definitely not the other way around), you first go from Oxford to London, then London to Cambridge. So the distance traveled is the distance from London to Oxford plus the distance from London to Cambridge. The exception is when the two destinations lie along the same line with London, in which case, you can directly take the train from one to the other without going through London, and hence the if $x = ky$ clause.

3.2 Norms and Inner Products

There are several notions on vector spaces that are closely related to metrics. We will look at norms and inner products of vector spaces, and show that they all naturally induce a metric on the space.

First of all, we define the norm. This can be thought of as the length of a vector in the vector space.

Definition 3.2. Let V be a real vector space. A *norm* on V is a function $\|\cdot\| : V \to \mathbb{R}$ such that

1. *Positive definite.* $\|v\| \geq 0$ for all $v \in V$, with $\|v\| = 0$ iff $v = 0$.
2. *Scaling.* For $\lambda \in \mathbb{R}$ and $v \in V$, $\|\lambda v\| = |\lambda|\|v\|$.
3. *Triangle inequality.* For $v, w \in V$, $\|v + w\| \leq \|v\| + \|w\|$.

A norm naturally induces a metric on V.

Lemma 3.3. If $\|\cdot\|$ is a norm on V, then

$$d(v, w) = \|v - w\|$$

defines a metric on V.

Proof.

(i) $d(v, w) = \|v - w\| \geq 0$ by definition.
$d(v, w) = 0 \iff \|v - w\| = 0 \iff v - w = 0 \iff v = w$.

(ii) $d(w, v) = \|w - v\| = \|(-1)(v - w)\| = |-1|\|v - w\| = d(v, w)$.

(iii) $d(u, v) + d(v, w) = \|u - v\| + \|v - w\| \geq \|u - w\| = d(u, w)$.

So d is indeed a metric. □

Examples.

(i) In $\mathbb{R}^n$ or $\mathbb{C}^n$, we have the *Euclidean norm*

$$\|x\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2}.$$

This induces our familiar Euclidean metric

$$d(x, y) = \|x - y\|_2 = \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2},$$

This is the standard metric on $\mathbb{R}^n$ or $\mathbb{C}^n$. The metric space (M, d) is called the n-dimensional real (or complex) *Euclidean space*, denoted ℓ_2^n. The Euclidean norm is denoted as the ℓ_2-norm, and the Euclidean metric is the ℓ_2-metric.

(ii) Again, in $\mathbb{R}^n$ or $\mathbb{C}^n$, we can have the *ℓ_p-norm* for $p \in [1, \infty)$ defined by

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p}.$$

This induces the *ℓ_p-metric*

$$d(x, y) = \left(\sum_{k=1}^{n} |x_k - y_k|^p \right)^{1/p}.$$

We can extend this to define the *ℓ_∞-norm* and the *ℓ_∞-metric*

$$\|x\|_\infty = \max_{1 \le k \le n} |x_k|,$$

$$d(x, y) = \max_{1 \le k \le n} |x_k - y_k|.$$

(iii) Let S be a set and let $L_\infty(S)$ be the set of all bounded scalar functions on S. We then define the *L_∞-norm* of $f \in L_\infty(S)$ by

$$\|f\|_\infty = \sup_{x \in S} |f(x)|.$$

This is known as the *sup norm* or the *uniform norm*. Then $d(f, g) = \|f - g\|_\infty$ defines a metric on $L_\infty(S)$ known as the *uniform metric*.

(iv) Consider $C[a, b]$, the set of continuous functions on $[a, b]$, then for $p \in [1, \infty)$, we define the *L_p-norm* and the *L_p-metric* as

$$\|f\|_p = \left(\int_a^b |f(x)|^p \right)^{1/p},$$

$$d(f, g) = \|f - g\|_p.$$

We will now define the inner product of a real vector space. This is a generalisation of the notion of the dot product.

Definition 3.4. Let V be a real vector space. An inner product is a function $\langle \cdot \mid \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies

1. *Positive definite.* $\langle v|v \rangle \ge 0$ for all $v \in V$, with $\langle v|v \rangle = 0$ iff $v = 0$.
2. *Symmetry.* $\langle v|w \rangle = \langle w|v \rangle$.
3. *Linearity.* $\forall \lambda \in \mathbb{R}$, $\langle u + \lambda v|w \rangle = \langle u|w \rangle + \lambda \langle v|w \rangle$.

Theorem 3.5 (Cauchy–Schwarz inequality). If $\langle \cdot \mid \cdot \rangle$ is an inner product, then

$$\langle v|w \rangle^2 \le \langle v|v \rangle \langle w|w \rangle .$$

Proof. For any x, we have

$$\langle v + xw|v + xw \rangle = \langle v|v \rangle + 2x \langle v|w \rangle + x^2 \langle w|w \rangle \ge 0.$$

This can be seen as a quadratic in x that is always non-negative. It can have at most one real root, so

$$(2 \langle v|w \rangle)^2 - 4 \langle v|v \rangle \langle w|w \rangle \le 0.$$

□

Lemma 3.6. If $\langle \cdot | \cdot \rangle$ is an inner product on V, then

$$\|v\| = \sqrt{\langle v|v \rangle}$$

is a norm.

Proof.

(i) $\|v\| = \sqrt{\langle v|v \rangle} \geq 0$.
$\|v\| = 0 \iff \langle v|v \rangle = 0 \iff v = 0$.

(ii) $\|\lambda v\| = \sqrt{\langle \lambda v|\lambda v \rangle} = \sqrt{\lambda^2 \langle v|v \rangle} = |\lambda|\|v\|$.

(iii) By Cauchy–Schwarz,

$$\begin{aligned}(\|v\| + \|w\|)^2 &= \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ &\geq \langle v|v \rangle + 2\langle v|w \rangle + \langle w|w \rangle \\ &= \|v + w\|^2 .\end{aligned}$$

□

Examples.

(i) Let $V = \mathbb{R}^n$. Then

$$\langle v|w \rangle = \sum_{i=1}^{n} v_i w_i$$

is an inner product. This is our familiar *dot product*, which induces the Euclidean norm and Euclidean metric.

(ii) Let $V = C[a, b]$. Then

$$\langle f|g \rangle = \int_a^b f(x)g(x)\,\mathrm{d}x$$

is an inner product. This induces the L_2-norm and the L_2-metric.

3.3 Subspaces and Product Spaces

Definition 3.7. Let (M, d) be a metric space and let $N \subset M$. Then d is a metric on N and (N, d) is called a *subspace* of M.

Example. Since every continuous function on a closed, bounded interval is bounded, so $C[a, b] \subset L_\infty([a, b])$. Therefore, $C[a, b]$ with the uniform metric is a subspace of $L_\infty([a, b])$.

Given two metric spaces, there is a number of natural ways to construct another metric space which has those two metric spaces as subspaces.

Definition 3.8. Let (M, d) and (N, d') be two metric spaces. We can define the metric *product space* by taking the set $M \times N$ and the metric

$$d_p((m_1, n_1), (m_2, n_2)) = (d(m_1, m_2)^p + d'(n_1, n_2)^p)^{1/p}$$

for some $p \in [1, \infty)$, or

$$d_\infty((m_1, n_1), (m_2, n_2)) = max\{d(m_1, m_2), d'(n_1, n_2)\} .$$

We denote the resulting product space $(M \times N, d_p)$ by $M \oplus_p N$.

Remark. There are no canonical choices on the value of p, but $p = 1, 2$ and ∞ are the most common. It is also straightforward to show that

$$d_\infty \leq d_2 \leq d_1 \leq 2d_\infty .$$

3.4 Convergence and Continuity

Metric spaces give us a measure of closeness. We can therefore extend our notion of convergence from $\mathbb{R}$ and $\mathbb{C}$ into general metric spaces.

Definition 3.9. Let (M, d) be a metric space. A sequence $a_n \in M$ is said to *converge* to a limit $a \in M$ if given any $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$, we have $d(a_n, a) < \epsilon$. We write $a_n \to a$ as $n \to \infty$.

Now we need to go back and check the theorems about convergence that we proved earlier to see whether they still hold for a general metric space.

Lemma 3.10 (Uniqueness of Limits). If $a_n \to a$ and $a_n \to b$ in a metric space (M, d), then $a = b$.

Proof. Assume that $a \neq b$. Given any $\epsilon > 0$, we can find integers N_1 and N_2 such that

$$\forall n \geq N_1, d(a_n, a) \leq \epsilon\,,$$

$$\forall n \geq N_2, d(a_n, b) \leq \epsilon\,.$$

Then letting $\epsilon = d(a, b)/3$ and taking $N = \max\{N_1, N_2\}$, we have, by the triangle inequality,

$$d(a, b) \leq d(a, a_n) + d(a_n, b) \leq 2\epsilon = \frac{2}{3} d(a, b)$$

for all $n \geq N$. Then we must have $d(a, b) = 0$ and $a = b$, a contradiction. □

Definition 3.11. Let $f : M \to M'$ be a function between metric spaces (M, d) and (M', d'). We say f is *continuous* at $a \in M$ if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\forall b \in M$ and $d(a, b) < \delta$, we have $d'(f(a), f(b)) < \epsilon$.

If f is continuous at every $a \in N \subseteq M$, then we say f is *continuous* on N.

Proposition 3.12 (Sequence definition of continuity). Let $f : M \to M'$ be a function between metric spaces, and let $a \in M$. Then f is continuous at a if and only if for any sequence $x_n \to a$, we have $f(x_n) \to f(a)$.

Proof.

($\Rightarrow$) If f is continuous at a, then $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$d(a, b) < \delta \implies d'(f(a), f(b)) < \epsilon\,.$$

Now, find any $x_n \to a$. We can find $N \in \mathbb{N}$ such that $\forall n > N$, $d(a, x_n) < \delta$. Therefore, for the same number of N, we have $\forall n > N$, $d(f(a), f(x_n)) < \epsilon$. So $f(x_n) \to f(a)$.

($\Leftarrow$) If $x_n \to a \implies f(x_n) \to f(a)$ but f is not continuous at a, then we can find $\epsilon > 0$ such that $\forall \delta > 0$, there exists some $x \in M$ such that $d(x, a) < \delta$ but $d'(f(x), f(a)) > \epsilon$. Set $\delta_n = \frac{1}{n}$ and obtain the corresponding x_n. Now $x_n \to a$ but $f(x_n) \not\to f(a)$. This is a contradiction. □

We then have the following two obvious corollaries.

Corollary. Let $f, g : M \to M'$ be continuous scalar functions, then $f + g$, $f \times g$ and f/g (providing that $\forall x \in M, g(x) \neq 0$) are all continuous.

Corollary. If $f : M \to M'$ and $g : M' \to M''$ are continuous, then $g \circ f$ is continuous.

Definition 3.13. A function $f : (M, d) \to (M, d')$ is *uniformly continuous* if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\forall x, y \in M$ with $d(x, y) < \delta$, we have $d'(f(x), f(y)) < \epsilon$.

Definition 3.14. A function $f : (M, d) \to (M', d')$ is *Lipschitz continuous* if there is some constant $C \geq 0$ such that $\forall x, y \in M$,

$$d'(f(x), f(y)) \leq Cd(x, y)\,.$$

We sometimes say f is *C-Lipschitz*.

Definition 3.15. A map $f : (N, d) \to (N', d')$ is *isometric* if $\forall x, y \in N$,

$$d'(f(x), f(y)) = d(x, y)\,.$$

Remark. Trivially, an isometric function is 1-Lipschitz and injective. If f is also surjective, then we call it an *isometry*. If there exists an isometry between two metric spaces, then we say that they are *isometric*.

Proposition 3.16. A Lipschitz function is uniformly continuous.

Proof. Trivial. □

3.5 Topology of a Metric Space

While we will discuss topology and topological spaces in general later, it's worth looking at a few of the core ideas in the context of metric spaces.

It seems that there are some concepts in metric spaces that don't really rely on the metric, just about the closeness of points. The point of this section is looking to see exactly what type of relations are these, so that we can generalise them to a setting where we don't use a metric.

Definition 3.17. Let (M, d) be a metric space. An *open ball* in M of center $x \in M$ and radius $r > 0$ is given by

$$D_r(x) \coloneqq \{y \in M \mid d(x, y) < r\}\,.$$

Then, we can say $x_n \to x$ in M if and only if for all $\epsilon > 0$, there is some integer N such that $n \geq N$ implies that $x_n \in D_\epsilon(x)$.

Similarly, for a function $f : M \to M'$, we say that f is continuous at $x \in M$ if and only if for all $\epsilon > 0$, there exists some $\delta > 0$ such that $f(D_\delta(x)) \subset D_\epsilon(f(x))$.

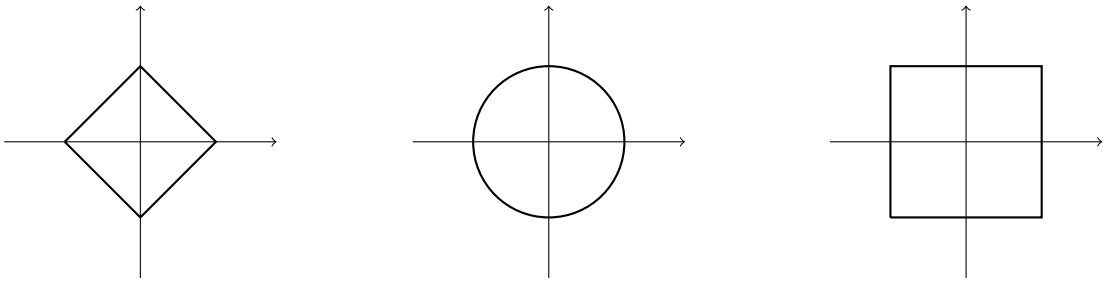
Definition 3.18. Let (M, d) be a metric space. A *closed ball* in M of a center $x \in M$ and radius $r > 0$ is given by

$$B_r(x) = \{y \in M \mid d(x, y) \leq r\}\,.$$

Examples.

(i) When $M = \mathbb{R}$, an open ball is an open interval and a closed ball is a closed interval.

(ii) In $\mathbb{R}^2$, a unit closed ball $B_1(0)$ has different appearance in different metrics.



d_1-metric d_2-metric d_∞-metric

(iii) If (M, d) is discrete, then $D_1(x) = \{x\}$, $B_1(x) = M$.

Note that $B_s(x) \subset D_r(x) \subset B_r(x)$ for any $s < r$.

Definition 3.19. A subset $U \subset M$ with $x \in U$ is called a *neighbourhood* of x if there exists some $r > 0$ such that $D_r(x) \subset U$.

Definition 3.20. Let $U \subset M$. We say U is *open* in M if for all $x \in U$, there exists some $r > 0$ such that $D_r(x) \subset U$, i.e. U is a neighbourhood of all of its points.

Example. Consider the upper half complex plane: $H = \{z \in \mathbb{C} \mid \operatorname{Im} z \geq 0\}$. Then let $w \in H$ and $\delta = \operatorname{Im} w$. If $\delta > 0$, then $D_\delta(w) \subset H$. If $\delta = 0$, then for all $r > 0$, $D_r(w) \not\subset H$. Thus H is not open.

Lemma 3.21. Open balls are open.

Proof. Consider an open ball $D_r(x) \subset M$. Let $y \in D_r(x)$ and let $\delta = r - d(x, y)$. Then $\delta > 0$, and if $z \in D_\delta(y)$ then

$$d(z, x) \leq d(z, y) + d(y, x) < \delta + d(x, y) = r\,,$$

so $z \in D_r(x)$. This shows that $D_\delta(y) \subset D_r(x)$. □

Corollary. Let M be a metric space, $U \subset M$ and $x \in M$. Then U is a neighbourhood of x if and only if there exists an open subset V of M such that $x \in V \subset U$.

3.5.1 Continuity and Convergence using Topology

Proposition 3.22 (Convergence using open sets). In a metric space M, the following are equivalent:

(i) $x_n \to x$;

(ii) for all neighbourhoods U of x in M, there exists some $N \in \mathbb{N}$ such that $\forall n \geq N$, $x_n \in U$.

(iii) for all open subsets U of M with $x \in U$, there exists some $N \in \mathbb{N}$ such that $\forall n \geq N$, $x_n \in U$.

Proof.

- (i) ⇒ (ii). Let U be a neighbourhood of x in M. Then by definition, $\exists \epsilon > 0$ such that $D_\epsilon(x) \in U$. Then since $x_n \to x$, there is an $N \in \mathbb{N}$ such that $n \geq N$ implies $d(x_n, x) < \epsilon$, and so $x_n \in D_\epsilon(x)$.
- (ii) ⇒ (iii). Trivial by the previous lemma.
- (iii) ⇒ (i). Given some $\epsilon > 0$, $U = D_\epsilon(x)$ is open and $x \in U$. Then by (iii) there is an $N \in \mathbb{N}$ such that $n \geq N$ implies $x_n \in U$, that is, $d(x_n, x) < \epsilon$. □

Proposition 3.23 (Local continuity using open sets). Let $f : M \to M'$ be a function between metric spaces, and let $x \in M$. The following are equivalent:

(i) f is continuous at x;

(ii) for all neighbourhood V of $f(x) \in M'$, there exists a neighbourhood U of x in M such that $f(U) \subset V$;

(iii) for all neighbourhoods V of $f(x)$ in M', $f^{-1}(V)$ is a neighbourhood of x in M.

Proof.

- (i) ⇒ (ii). Let V be a neighbourhood of $f(x)$ in M'. Then by definition there exists some $\epsilon > 0$ such that $D_\epsilon(f(x)) \subset V$. Since f is continuous at x, there exists $\delta > 0$ such that $f(D_\delta(x)) \subset D_\epsilon(f(x))$. Then $U = D_\delta(x)$ is a neighbourhood of x in M and $f(U) \subset V$.

- (ii) $\Rightarrow$ (iii). Let V be a neighbourhood of $f(x)$ in M'. Then by (ii) there is a neighbourhood U of x in M such that $f(U) \subset V$. Then $U \subset f^{-1}(V)$, and since U is a neighbourhood of x in M, there is some $r > 0$ with $D_r(x) \subset U \subset f^{-1}(V)$. Thus $f^{-1}(V)$ is a neighbourhood of x in M.
- (iii) $\Rightarrow$ (i). Given $\epsilon > 0$, $V = D_\epsilon(f(x))$ is a neighbourhood of $f(x)$ in M. By (iii), $f^{-1}(V)$ is a neighbourhood of x in M. So there exists $\delta > 0$ with $D_\delta(x) \subset f^{-1}(V)$. Then $f(D_\delta(x)) \subset V = D_\epsilon(f(x))$. □

Proposition 3.24 (Global continuity using open sets). Let $f : M \to M'$ be a function between metric spaces. Then

$$f \text{ is continuous } \iff f^{-1}(V) \text{ is open in } M \text{ for all open subsets } V \text{ of } M'.$$

Proof.

($\Rightarrow$) Let V be open in M'. Let $x \in f^{-1}(V)$. Then $f(x) \in V$. Since V is open, there is some $\epsilon > 0$ such that $D_\epsilon(f(x)) \subset V$. Since f is continuous at x, there is a $\delta > 0$ such that $f(D_\delta(x)) \subset D_\epsilon(f(x))$. Therefore $D_\delta(x) \subset f^{-1}(D_\epsilon(f(x))) \subset f^{-1}(V)$. Thus $f^{-1}(V)$ is open in M.

($\Leftarrow$) Let $x \in M$ and let $\epsilon > 0$. Then $V = D_\epsilon(f(x))$ is open in M'. By (ii), $f^{-1}(V)$ is open in M, also $x \in f^{-1}(V)$ and $f(x) \in V$. Then by definition there is a $\delta > 0$ such that $D_\delta(x) \subset f^{-1}(V)$, so $f(D_\delta(x)) \subset V = D_\epsilon(f(x))$. □

3.5.2 Topology of a Metric space

We can now write down what the topology of a metric space is. We can use some properties that occur here to inform how we define things more generally afterwards.

Definition 3.25. The *topology of a metric space* (M, d) is the family of all open subsets of M.

Proposition 3.26 (Topology of metric spaces). The topology of a metric space satisfies the following properties:

(i) $\emptyset$ and M are open;

(ii) if U_i are open in M for $i \in I$ (the index set I may be countable or uncountable), then $\bigcup_{i\in I} U_i$ is open in M;

(iii) if U, V are open then $U \cap V$ is open.

Proof.

(i) This is clear.

(ii) Given $x \in \bigcup_{i\in I} U_i$, then there is some $i_0 \in I$ with $x \in U_{i_0}$. Then U_{i_0} is open, so by definition there is an $r > 0$ such that $D_r(x) \subset U_{i_0} \subset \bigcup_{i\in I} U_i$.

(iii) Given $x \in U \cap V$, since U is open and $x \in U$, there is an $r > 0$ with $D_r \subset U$, and since V is open and $x \in V$, there is an $s > 0$ with $D_s(x) \subset V$. Set $t = \min\{r, s\}$, we then have $t > 0$ and $D_t(x) = D_r(x) \cap D_s(x) \subset U \cap V$. □

Definition 3.27. A subset A of a metric space (M, d) is *closed* in M if for every sequence $x_n \in A$ that is convergent in M, we have $\lim_{n\to\infty} x_n \in A$.

Lemma 3.28. Closed balls are closed.

Proof. Consider $B_r(x) = \{y \in M \mid d(x, y) \leq r\} \subset M$, and a sequence $x_n \in B_r(x)$ such that $x_n \to z$. We need to show that $z \in B_r(x)$. We have

$$\begin{aligned} d(z, x) &\leq d(z, x_n) + d(x_n, x) \\ &\leq d(z, x_n) + r \to r \end{aligned}$$

as $n \to \infty$. Thus $d(z, x) \leq r$ and hence $z \in B_r(x)$. □

Examples.

(i) The set $[0,1] = B_{1/2}(1/2)$ is closed in $\mathbb{R}$. It is not open as $D_r(0) \not\subseteq [0,1]$ for any $r > 0$.

(ii) The set $(0,1) = D_{1/2}(1/2)$ is open in $\mathbb{R}$, and it is not closed. Take $a_n = 1/(n+1) \in (0,1)$ for all $n \in \mathbb{N}$, but $a_n \to 0 \notin (0,1)$.

(iii) $\mathbb{R}$ and $\emptyset$ are open and closed in $\mathbb{R}$.

(iv) The set $(0,1]$ in $\mathbb{R}$ is neither open nor closed.

(v) If (M,d) is discrete, then every subset of M is both open and closed.

Lemma 3.29. Let $A \subseteq M$. Then A is closed in M if and only if $M \setminus A$ is open in M.

Proof. Assume A is closed and $M \setminus A$ is not open. Then there exists $x \in M \setminus A$ such that for all $r > 0$ we have $D_r(x) \not\subset M \setminus A$. Hence for all n, there exists $x_n \in D_{1/n}(x) \cap A$. Then $d(x_n, x) < 1/n \to 0$, so $x_n \to x \notin A$, but $x_n \in A$ for all n, which contradicts that A is closed.

Suppose $M \setminus A$ is open but A is not closed. Then there exists some sequence $x_n \in A$ such that $x_n \to x \in M$ but $x \notin A$. Since $x \in M \setminus A$ and $M \setminus A$ is open, there exists $\epsilon > 0$ with $D_\epsilon(x) \subset M \setminus A$. Then since $x_n \to x$, there is some $N \in \mathbb{N}$ such that for all $n \geq N$, we have $x_n \in D_\epsilon(x)$, and hence $x_n \in M \setminus A$, which is a contradiction. □

We previously met isometries, which were maps between metric spaces that preserved the distance function. Now that we are trying to look slightly away from metrics and more towards open sets and neighbourhoods, it's natural to think of a new definition of a function between metric spaces that preserves open sets.

Definition 3.30. A map $f : M \to M'$ between metric spaces is a *homeomorphism* if f is a bijection and f and f^{-1} are both continuous.

Equivalently, f is a *homeomorphism* if for all open sets $V \subset M'$, $f^{-1}(V)$ is open in M, and for all open sets $U \subset M$, $f(U)$ is open in M'.

If there is a homeomorphism between M and M', we say that M and M' are *homeomorphic*.

Example. The metric spaces $(0,\infty)$ and $(0,1)$ are homeomorphic with the homeomorphism $x \mapsto 1/(x+1)$ and inverse $x \mapsto 1/x - 1$.

Remark. It is true that every isometry is a homeomorphism, but the converse is clearly false.

Definition 3.31. Let d and d' be metrics on a set M. We say that d and d' are *equivalent*, denoted $d \sim d'$ if they define the same topology on M, i.e. for $U \subset M$, U is open in (M,d) if and only if U is open in (M,d').

Remark. $d \sim d'$ if and only if $\mathrm{id} : (M,d) \to (M,d')$ is a homeomorphism.

Also note that if $d \sim d'$, then (M,d) and (M,d') have the same convergent sequences and the same continuous maps.

Definition 3.32. Let d and d' be metrics on a set M. We say that d and d' are *uniformly equivalent* if $\mathrm{id} : (M,d) \to (M,d')$ and $\mathrm{id}' : (M,d') \to (M,d)$ are uniformly continuous, written $d \sim_u d'$.

We say d and d' are *Lipschitz equivalent* if these functions are Lipschitz, written $d \sim_{Lip} d'$.

Remarks.

(i) $d \sim_{Lip} d'$ if and only if there exists $a > 0$ and $b > 0$ such that

$$ad(x,y) \leq d'(x,y) \leq bd(x,y)\,.$$

(ii)
$$d \sim_{Lip} d' \implies d \sim_u d' \implies d \sim d' .$$

Examples.

(i) Given a metric space (M, d), $d'(x, y) = \min\{1, d(x, y)\}$ defines a metric on M, and $d' \sim_u d$.

(ii) On a product space $M \times M'$, the metrics d_1, d_2 and d_∞ are pairwise Lipschitz equivalent.

(iii) On $C[0, 1]$, the L_1-metric and the uniform metric are not equivalent.

3.6 Completeness

In $\mathbb{R}$ and $\mathbb{C}$, every Cauchy sequence is convergent. We wish to generalise this notion to an arbitrary metric space.

Definition 3.33. A sequence x_n in a metric space (M, d) is *Cauchy* if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall m, n \geq N$, we have $d(x_m, x_n) < \epsilon$.

Definition 3.34. A sequence x_n in a metric space (M, d) is *bounded* if $\exists z \in M$ and $r > 0$ such that $\forall n \in \mathbb{N}$, we have $x_n \in B_r(z)$.

Lemma 3.35. In a metric space (M, d), every convergent sequence is Cauchy, and every Cauchy sequence is bounded.

Proof. Let $x_n \in M$ be convergent, and let $x = \lim_{n\to\infty} x_n$. Given $\epsilon > 0$, there is some positive integer N such that $d(x_n, x) \leq \epsilon/2$ for all $n \geq N$. Then for all $m, n \geq N$, $d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \epsilon$. Thus a convergent sequence is Cauchy.

Now assume x_n is Cauchy in M. Then there is a positive integer N such that for all $m, n \geq N$, we have $d(x_m, x_n) < 1$. So in particular, we have $d(x_n, x_N) < 1$ for all $n \geq N$. Let $r = \max\{d(x_1, x_n), \ldots, d(x_{N-1}, x_N), 1\}$. Then $x_n \in B_r(x_N)$ for all $n \in N$, and so x_n is bounded. □

In general, a bounded sequence is not Cauchy. An example is the oscillating sequence $0, 1, 0, 1 \ldots$ in $\mathbb{R}$.

A more subtle observation is that a Cauchy sequence is not necessarily convergent in general metric space. For example, $x_n = 1/n$ is clearly Cauchy, but it does not converge in $(0, \infty)$ equipped with the usual metric.

It seems that this example works because there is something 'missing' from the metric space.

Definition 3.36. A metric space (M, d) is *complete* if every Cauchy sequence in M converges in M.

Trivial examples of complete metric spaces would be $\mathbb{R}$ and $\mathbb{C}$.

Proposition 3.37 (Product of complete spaces is complete). If metric spaces (M, d) and (M', d') are complete, then so is $M \oplus_p M'$.

Proof. Let a_n be a Cauchy sequence in the product space $M \oplus_p M'$. We will write $a_n = (x_n, x'_n)$, where x_n and x'_n are sequences in M and M' respectively. As a_n is Cauchy, $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall m, n \geq N$, we have $d_p(a_m, a_n) < \epsilon$. Then, $\forall m, n \geq N$,
$$d(x_m, x_n) \leq \max\{d(x_m, x_n), d'(x'_m, x'_n)\} \leq d_p(a_m, a_n) < \epsilon .$$
Hence, x_n is Cauchy in M, and similarly x'_n is Cauchy in M'. Since M and M' are complete, x_n and x'_n converges to $x \in M$ and $x' \in M'$ respectively. Let $a = (x, x')$, then
$$d_p(a_n, a) \leq d_1(a_n, a) = d(x_n, x) + d(x'_n, x') \to 0 ,$$
so the product space is complete. □

Corollary. $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete in ℓ_p-metric.

Theorem 3.38 (Completeness of function spaces). Let S be any non-empty set. Then $L_\infty(S)$, the set of all bounded scalar functions on S, is complete in the uniform metric D.

Proof. Let f_n be a Cauchy sequence in $L_\infty(S)$. Then given $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n, m \geq N$ we have

$$D(f_m, f_n) = \sup_{x \in S} |f_m(x) - f_n(x)| < \epsilon \,.$$

So f_n is uniformly Cauchy. By the general principle of uniform convergence, f_n converges uniformly to some function f on S. We know that f is bounded, so $f \in L_\infty(S)$. Now it remains to show that $f_n \to f$ in the uniform metric. We have that $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$ and $x \in S$, $|f_n(x) - f(x)| < \epsilon$. This translates directly to

$$\sup_{x \in S} |f_n(x) - f(x)| = D(f_n, f) \leq \epsilon$$

for all $n \geq N$, which is convergence in the uniform metric as required. □

Proposition 3.39 (Completeness of a subspace). Let N be a subspace of a metric space M. Then

(i) if N is complete, N is closed in M.

(ii) If M is complete and N is closed in M, then N is complete.

Hence in a complete metric space, a subspace is complete if and only if it is closed.

Proof.

(i) Trivial. Any convergent sequence $x_n \in N$ is Cauchy in N, and so converges to some point $x \in N$ by completeness. It is therefore closed.

(ii) Let x_n be a Cauchy sequence in N. Then x_n is Cauchy in M, so $x_n \to x \in M$ by the completeness if M. Since N is closed in M, we must have $x \in N$, so $x_n \to x$ in N. □

Theorem 3.40. Let (M, d) be a metric space, and let

$$C_b(M) = \{f \in L_\infty(M) \mid f \text{ is continuous}\} \,.$$

This is a subspace of $L_\infty(M)$ in the uniform metric D.

$C_b(M)$ is complete in the uniform metric.

Proof. By the previous proposition, it suffices to show that $C_b(M)$ is closed in $L_\infty(M)$. Let $f_n \in C_b(M)$ be a sequence, and assume that $f_n \to f$ in $L_\infty(M)$. We need to show that f is continuous.

Given some $a \in M$ and $\epsilon > 0$, we can fix an $n \in \mathbb{N}$ such that $D(f_n, f) < \epsilon/3$. Since f_n is continuous at a, there is a $\delta > 0$ such that for all $x \in M$, $d(x, a) < \delta$, we have $|f_n(x) - f_n(a)| < \epsilon$. Hence for all $x \in M$, if $d(x, a) < \delta$ then

$$\begin{aligned} |f(x) - f(a)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| \\ &\leq 2D(f_n, f) + |f_n(x) - f_n(a)| \\ &\leq \epsilon \,. \end{aligned}$$

□

Corollary. $C[a, b]$ is complete in the uniform metric.

Proof. $C[a, b] = C_b[a, b]$. □

Definition 3.41. Let S be a non-empty set and (N, d') be a metric space. Define the space of functions

$$L_\infty(S, N) \coloneqq \{f : S \to N \mid f \text{ is bounded}\}.$$

For $f, g \in L_\infty(S, N)$, as they are bounded, we have

$$f(x) \in B_r(y)\,,\ g(x) \in B_s(z)$$

for all $x \in S$ and some $y, z \in N$. Then $\forall x \in S$, we have

$$d'(f(x), g(x)) \le d'(f(x), y) + d'(y, z) + d'(z, g(x)) \le r + s + d'(y, z)\,.$$

This is a uniform bound for all x, so we may take the supremum.

Definition 3.42. The *uniform metric* on $L_\infty(S, N)$, where S is a non-empty set and (N, d') is a metric space, is defined as

$$D(f, g) \coloneqq \sup_{x \in S} d'(f(x), g(x))$$

for $f, g \in L_\infty(S, N)$.

Definition 3.43. Let (M, d) and (N, d') be metric spaces. Define

$$C_b(M, N) = \{f : M \to N \mid f \text{ continuous and bounded}\}.$$

Remark. $C_b(M, N)$ is a subspace of $L_\infty(M, N)$ with the uniform metric.

Theorem 3.44. Let S be a set, let (M, d) be a metric space and let (N, d') be a complete metric space. Then

(i) $L_\infty(S, N)$ is complete in the uniform metric D;

(ii) $C_b(M, N)$ is complete in the uniform metric D.

Proof.

(i) Let $f_n \in L_\infty(S, N)$ be a Cauchy sequence. We first show that f_n is pointwise Cauchy. Let $x \in S$. Then $\forall \epsilon > 0$, $\exists K \in \mathbb{N}$ such that $\forall m, n \ge K$, $D(f_m, f_n) < \epsilon$. In particular, $d'(f_m(x), f_n(x)) \le D(f_m, f_n) < \epsilon$. Therefore, the sequence $f_k(x)$ is Cauchy in N $\forall x \in S$.

Since N is complete, $f_k(x)$ converges in N for all $x \in S$. Hence, we can define the pointwise limit of $f_n(x)$ as $f : S \to N$, $f(x) = \lim_{n\to\infty} f_k(x)$.

Now we have to show that f is bounded. Since f_k is Cauchy in the uniform metric D, there exists $K \in \mathbb{N}$ such that $\forall m, n \ge K$, $D(f_i, f_j) < 1$. Then in particular, for all $n > K$, $D(f_n, f_K) < 1$. Then since f_K is bounded, $\exists y \in N$, $r > 0$ such that $\forall x \in S$, $f_K(x) \in B_r(y)$. Then by the triangle inequality, $d'(f_n(x), f_K(x)) \le 1$. Hence

$$d'(f(x), y) \le d'(f(x), f_K(x)) + d'(f_K(x), y) \le 1 + r$$

for all x. This is a uniform bound, so $f \in L_\infty(S, N)$ and hence $L_\infty(S, N)$ is complete.

In addition, we can show that $f_k \to f$ uniformly in D. We have $\forall \epsilon > 0$, $\exists K \in \mathbb{N}$ such that $\forall m, n \ge K$, $D(f_m, f_n) < \epsilon$. Choose $n \ge K$ and $x \in S$. Then for all $m \ge K$, $d'(f_m(x), f_n(x)) \le D(f_m, f_n) < \epsilon$. As $m \to \infty$, $d'(f(x), f_n(x)) \le \epsilon$ because the metrics are continuous. Since x is arbitrary and the uniform distance $D(f, f_n) < \epsilon$ holds for all $n \ge K$, we have uniform convergence.

(ii) By (i), it is enough to show that $C_b(M, N)$ is closed in $L_\infty(M, N)$. Let $f_k \in C_b(M, N)$ and $f_k \to f \in L_\infty(M, N)$. We need to show that f is continuous. This is exactly the proof that the uniform limit of continuous functions is continuous. Let $a \in M$, $\epsilon > 0$. Then as $f_k \to f \in L_\infty(M, N)$, we can fix $k \in \mathbb{N}$ such that $D(f_k, f) < \epsilon$. Since f_k is continuous, $\exists \delta > 0$ such that $\forall x \in M$, $d(x, a) < \delta$, we have $d'(f_k(x), f_k(a)) < \epsilon$, so

$$\begin{aligned} d'(f(x), f(a)) &\le d'(f(x), f_k(x)) + d'(f_k(x), f_k(a)) + d'(f_k(a), f(a)) \\ &< 3\epsilon\,. \end{aligned}$$

□

3.7 Contraction Mapping Theorem

3.7.1 Contraction Mapping Theorem

Definition 3.45. A function $f : M \to M'$ is called a *contraction mapping* if $\exists \lambda \in [0, 1)$ such that $\forall x, y \in M$,

$$d'(f(x), f(y)) \leq \lambda d(x, y)\,,$$

so that f is λ-Lipschitz.

We have the contraction mapping theorem, also known as the Banach's fixed point theorem.

Theorem 3.46 (Contraction mapping theorem). Let M be a non-empty complete metric space. Let $f : M \to M$ be a contraction mapping. Then f has a unique fixed point:

$$\exists! z \in M\,,\ f(z) = z\,.$$

Proof. First we can show the uniqueness easily. Suppose there are two fixed points $f(z) = z$, $f(w) = w$. Then $d(z, w) = d(f(z), f(w)) \leq \lambda d(z, w) < d(z, w)$. Hence $d(z, w) = 0$ so $z = w$.

Now we show the existence. Fix a starting point $x_0 \in M$. Let $x_n = f(x_{n-1})$ for all $n \in \mathbb{N}$, so $x_n = f^n(x_0)$. First, observe that for all $n \in \mathbb{N}$,

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq \lambda d(x_{n-1}, x_n) \leq \cdots \leq \lambda^n d(x_0, x_1)\,.$$

For $m \geq n$, we have

$$d(x_n, x_m) \leq \sum_{k=n}^{m-1} d(x_k, x_{k+1}) \leq \sum_{k=1}^{m-1} \lambda^k d(x_0, x_1) \leq \frac{\lambda^n}{1-\lambda} d(x_0, x_1)\,.$$

Since $\lambda^n/(1-\lambda)d(x_0, x_1) \to 0$, $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$,

$$\frac{\lambda^n}{1-\lambda} d(x_0, x_1) < \epsilon\,.$$

Hence, $\forall m \geq n \geq N$, $d(x_n, x_m) \leq \epsilon$, so the sequence x_n is Cauchy. Since M is complete, x_n is convergent to some point $z \in M$. f is continuous since it is λ-Lipschitz, so as $f(x_n) \to z$, $f(z) = z$. The fixed point exists. □

Remark. Let $m \to \infty$ in the inequality for $d(x_n, x_m)$, get

$$d(x_n, z) \leq \frac{\lambda^n}{1-\lambda} d(x_0, x_1)\,.$$

So $x_0 \to z$ exponentially fast.

Non-examples.

(i) Consider

$$f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}\,,$$
$$x \mapsto \frac{x}{2}\,.$$

This is a contraction, but there is no fixed point as $\mathbb{R} \setminus \{0\}$ is not complete.

(ii) Consider

$$f : [1, \infty) \to [1, \infty) ,$$
$$x \mapsto x + \frac{1}{x} .$$

Certainly $|f(x) - f(y)| < |x - y|$, and $[1, \infty)$ is closed in $\mathbb{R}$ so it is complete. However, this is not a contraction: even though $|f(x) - f(y)| < |x - y|$, there is no upper bound λ (the supremum is exactly 1). There is no fixed point.

Example. Let $y_0 \in \mathbb{R}$. Then the initial value problem

$$\begin{cases} f'(t) = f(t^2) \\ f(0) = y_0 \end{cases}$$

has a unique solution on $C\big[0, \frac{1}{2}\big]$.

Assume that there is a solution, then immediately f is continuously differentiable. By the fundamental theorem of calculus,

$$f(t) = f(0) + \int_0^t f(x^2)\,\mathrm{d}x \ .$$

Let $M = C\big[0, \frac{1}{2}\big]$ with the uniform metric, which is non-empty and complete. Consider the mapping $T : M \to M$ defined by

$$(Tg)(t) = y_0 + \int_0^t g(s^2)\,\mathrm{d}s \ .$$

Note that Tg is well-defined since $g(s^2)$ is continuous. Moreover, by the fundamental theorem of calculus, Tg is differentiable and $(Tg)'(t) = g(t^2)$. Thus f is a solution to the initial value problem if and only if $f \in M$ and $Tf = f$. Now, if T is a contraction, then we can assert that there is a unique solution.

For $g, h \in M$, $t \in \big[0, \frac{1}{2}\big]$, consider

$$|(Tg)(t) - (Th)(t)| = \left| \int_0^t g(s^2) - h(s^2)\,\mathrm{d}s \right| \leq \int_0^t \big|g(s^2) - h(s^2)\big|\,\mathrm{d}s \leq \frac{1}{2} D(g, h) .$$

Taking the supremum over t gives $D(Tg, Th) \leq \frac{1}{2} D(g, h)$, and so there is exactly one fixed point.

Remark. The above example shows that for any $\delta \in (0, 1)$, there is a unique solution to the initial value problem on $[0, \delta]$ called f_δ, since $\delta < 1$ is required for the map to be a contraction. For $0 < \delta < \mu < 1$, $f_\mu|_{[0,\delta]} = f_\delta$ by uniqueness. So we can combine the solution together to yield a unique solution on $[0, 1)$.

3.7.2 Lindelöf–Picard Theorem

Theorem 3.47 (Lindelöf–Picard theorem). Let $n \in \mathbb{N}$, $y_0 \in \mathbb{R}^n$ and $a, b, R \in \mathbb{R}$ such that $a < b$ and $R > 0$. Let $\phi : [a, b] \times B_R(y_0) \to \mathbb{R}^n$ be a continuous function. Given that there exists $K > 0$ such that $\forall t \in [a, b]$ and $x, y \in B_R(y_0)$,

$$\|\phi(t, x) - \phi(t, y)\| \leq K \|x - y\| .$$

Then $\exists \epsilon > 0$ such that $\forall t, t_0 \in [a, b]$, the initial value problem

$$\begin{cases} f'(t) = \phi(t, f(t)) \\ f(t_0) = y_0 \end{cases}$$

has a unique solution on $[c, d] = [t_0 - \epsilon, t_0 + \epsilon] \cap [a, b]$.

Remark. Note that if $f : [c, d] \to \mathbb{R}^n$, we let $f_k : [c, d] \to \mathbb{R}$ to be the k-th component of f. We define $f'(t) \coloneqq (f_1'(t), \dots, f_n'(t))$ and

$$\int_c^d f(t)\,\mathrm{d}t \coloneqq \left(\int_c^d f_1(t)\,\mathrm{d}t\,, \dots, \int_c^d f_n(t)\,\mathrm{d}t\right).$$

Note that we can use the Cauchy–Schwarz inequality to give

$$\left\|\int_c^d f(t)\,\mathrm{d}t\right\| \le \int_c^d \|f(t)\|\,\mathrm{d}t \le (d-c)\sup_{t\in[c,d]}\|f(t)\|\,.$$

Proof. Since ϕ is a continuous function on $[a, b] \times B_R(y_0)$, which is closed and bounded, $\|\phi\|$ is bounded above by some $C > 0$. Let $\epsilon = \min\{\frac{R}{C}, \frac{1}{2K}\}$. Let $t_0 \in [a, b]$ and $[c, d] = [t_0 - \epsilon, t_0 + \epsilon] \cap [a, b]$. We need to show that there exists a unique differentiable function $f : [c, d] \to \mathbb{R}^n$ such that $f(t_0) = y_0$ and $f'(t) = \phi(t, f(t))$ for all $t \in [c, d]$.

Since $B_R(y_0)$ is closed in the complete space $\mathbb{R}^n$, it is also complete. Then $M = C([c, d], B_R(y_0))$ is non-empty and complete in the uniform metric D. A natural thought is then to reduce this IVP to a fixed-point problem. By applying the fundamental theorem of calculus coordinate-wise, f is a solution to the initial value problem if $f \in M$ and

$$f(t) = y_0 + \int_{t_0}^t \phi(s, f(s))\,\mathrm{d}s\ .$$

Consider the mapping $T : M \to M$ given by

$$(Tg)(t) = y_0 + \int_{t_0}^t \phi(x, g(x))\,\mathrm{d}x\ .$$

We must first show that T is well-defined. Note that the integral is well defined: $s \to \phi(s, g(s))$ is continuous so integrable. By the fundamental theorem of calculus, Tg is differentiable and the derivative is $(Tg)'(t) = \phi(t, g(t))$. In particular, $Tg : [c, d] \to \mathbb{R}^n$ is continuous. Finally, for $t \in [c, d]$,

$$\|(Tg)(t) - y_0\| = \left\|\int_{t_0}^t \phi(s, g(s))\,\mathrm{d}s\right\| \le |t - t_0| \sup_{s\in[c,d]}\|\phi(s, g(s))\| \le \epsilon C \le R\,,$$

so $Tg \in M$.

Recall that f is a solution to the IVP if and only if $f \in M$ and $Tf = f$. We will show that T is a contraction mapping so it has a unique fixed point. Let $t \in [c, d]$ and $g, h \in M$.

$$\|(Tg)(t) - (Th)(t)\| = \left\|\int_{t_0}^t \phi(s, g(s)) - \phi(s, h(s))\,\mathrm{d}s\right\|.$$

Note that $\|\phi(s, g(s)) - \phi(s, h(s))\| \le K\|g(s) - h(s)\| \le KD(g, h)$, so

$$\|(Tg)(t) - (Th)(t)\| \le |t - t_0|KD(g, h) \le \epsilon KD(g, h)\,.$$

Taking the supremum over $t \in (c, d)$,

$$D(Tg, Th) \le \epsilon KD(g, h) \le \frac{1}{2}D(g, h)\,,$$

so T is a contraction. Done by contraction mapping theorem. □

Remark. In general, we cannot extend the solution guaranteed above to a global solution. However, we can apply the theorem to ODE of any order.

4 Topological Spaces

4.1 Topological Spaces

Definition 4.1. Let X be a set. A *topology* on X is a collection τ of subsets of X such that the following axioms hold:

1. $\emptyset, X \in \tau$;
2. if $U_i \in \tau$ for all $i \in I$, where I is some index set, then $\bigcup_{i\in I} U_i \in \tau$;
3. if $U, V \in \tau$, then $U \cap V \in \tau$.

A *topological space* is a pair (X, τ), where τ is a topology on a set X. Elements in τ are called the *open sets*.

Remark. If $U_i \in \tau$ for $i = 1, \dots, n$, then

$$\bigcap_{i=1}^{n} U_i \in \tau\,.$$

Definition 4.2. Let (M, d) be a metric space. Then the topology *induced* by d is the set of all open sets of M under d.

Remark. For a set open specifically in the topological sense, we say that it is τ-open. For a metric space (M, d), a set $U \subset M$ is open in metric sense if $\forall x \in U$, $\exists r > 0$ such that $D_r(x) \subset U$. We may say that U is d-open. We have proven that the collection of d-open sets in a metric space (M, d) is a topology τ on M.

Definition 4.3. Let (X, τ) be a topological space. We say that X is *metrizable* if there exists a metric d on X such that τ is the metric topology on X induced by d.

Remark. If $d' \sim d$, then d' and d induce the same topology on X.

Example. It is easy to show that $d_1 \sim d_2 \sim d_\infty$. All d_p metrics induce the same topology on $\mathbb{R}^n$.

Definition 4.4. Let τ_1, τ_2 be topologies on X. We say that τ_1 is *coarser* than τ_2, or τ_2 is *finer* than τ_1, if $\tau_1 \subset \tau_2$.

Examples.

(i) The *indiscrete topology* on a set X is the topology $\tau = \{\emptyset, X\}$. It is the coarsest topology on X.

If $|X| \geq 2$, then this is not metrizable. Let d be a metric on X, $x \neq y \in X$, $r = d(x, y)$ and $U = D_r(x)$. We know that U is d-open. But since $x \in U$, $y \notin U$, $U \notin \tau$.

(ii) The *discrete topology* on a set X is $\tau = \mathcal{P}(X)$, the power set of X. This is the finest topology on X. It is metrizable by the discrete metric.

Definition 4.5. A topological space is *Hausdorff* if $\forall x \neq y \in X$, there exists open sets $U, V \subset X$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

Informally, x, y are separated by open sets.

Proposition 4.6. Metric spaces are Hausdorff.

Proof. Let $x \neq y$ be points in a metric space (M, d). Let $r > 0$ such that $2r < d(x, y)$. Then let $U = D_r(x)$ and $V = D_r(y)$. U and V are certainly open, and they have no intersection by the triangle inequality, so the metric space is Hausdorff. □

Example. The *cofinite topology* on a set X is

$$\tau = \{\emptyset\} \cup \{U \subset X \mid U \text{ is cofinite in } X\},$$

where U is *cofinite* in X if $X \setminus U$ is finite.

When X is finite, this topology is simply the discrete topology $\tau = \mathcal{P}(X)$. When X is infinite, τ is not metrizable. Let $x \neq y \in X$, and let $x \in U$, $y \in V$, where U, V are open in X. Then U, V are cofinite, and hence $U \cap V \neq \emptyset$. So this topology on an infinite set is not Hausdorff and hence not metrizable.

Definition 4.7. A subset A of a topological space (X, τ) is *closed* in X if $X \setminus A$ is open in X.

Remark. In a metric space, this agrees with the earlier definition of a closed subset, as proven above.

Proposition 4.8. The collection of closed sets in a topological space X satisfies:

(i) $\emptyset, X$ are closed;

(ii) if A_i are closed in X for $i \in I$, where I is some non-empty index set, then $\bigcap_{i \in I} A_i$ is closed;

(iii) if A_1, A_2 are closed in X, then $A_1 \cup A_2$ is closed.

Examples.

(i) In a discrete topological space, every set is closed.

(ii) In the cofinite topology, a subset is closed if and only if it is finite or the full set.

4.2 Neighbourhoods and Convergence

Definition 4.9. Let (X, τ) be a topological space, and let $U \subset X$ and $x \in X$. We say that U is a *neighbourhood* of x in X if there exists an open set $V \in \tau$ such that $x \in V \subset U$.

Remark. Again, this agrees with our definition in a metric space.

Proposition 4.10. Let (X, τ) be a topological space and $U \subset X$. Then U is open if and only if U is a neighbourhood of x for every $x \in U$.

Proof. If U is open and $x \in U$, then let $V = U$, so that V is open and $x \in V \subset U$.

Conversely, if U is a neighbourhood of x, there exists $V_x \subset X$ such that $x \in V_x \subset U$. Then $U = \bigcup_{x \in U} V_x$ is open, since each V_x is open. □

Definition 4.11. Let $x_n \in X$ be a sequence in a topological space (X, τ). Let $x \in X$. We say that x_n *converge* to x if for all neighbourhoods U of x, there exists $N \in \mathbb{N}$ such that $\forall n \geq N$, $x_n \in U$.

Equivalently, x_n converge to x if for all open sets U containing x, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $x_n \in U$.

Remark. Again, the definition in a metric space agrees with this definition.

Examples.

(i) Eventually constant sequences converge.
If $\exists z \in X$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $x_n = z$, then $x_n \to z$.

(ii) In an indiscrete topological space, every sequence converges to every point.

(iii) In the cofinite topology on a set X, suppose that $x_n \to x \in X$. Then if $y \neq x$, $X \setminus \{y\}$ is a neighbourhood of x. Then $N_y = \{n \in \mathbb{N} \mid x_n = y\}$ is finite.

Conversely, suppose x_n is a sequence such that for some $x \in X$ and all $y \neq x$, N_y is finite, then $x_n \to x$.

In particular, if N_y is finite for all $y \in X$, the sequence converges to every point.

Proposition 4.12 (Convergence in a Hausdorff space). If $x_n \to x$ and $x_n \to y$ in a Hausdorff space, then $x = y$.

Proof. Suppose $x \neq y$, then we can choose open sets U and V such that $x \in U, y \in V$ with $U \cap V = \emptyset$. Since $x_n \to x$, there exists $N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$, $x_n \in U$. Similarly, there is a $N_2 \in \mathbb{N}$ such that $\forall n \geq N_2$, $x_n \in V$. Hence $\forall n \geq \max\{N_1, N_2\}$, $x_n \in U$ and $x_n \in V$. But $U \cap V = \emptyset$, so this is a contradiction. □

Remark. In a metric space, a subset A is closed if and only if whenever $x_n \in A$ and $x_n \to x$ implies $x \in A$. In a general topological space, any closed set is closed under limits, but not every subset that is closed under limits is closed.

4.3 Interior and Closure

Definition 4.13. Let (X, τ) be a topological space, and $A \subset X$. We define the *interior* of A, denoted A° or $\text{int}(A)$, by

$$A^\circ := \bigcup_{U \subset A,\, U \text{ open}} U\,.$$

The *closure* of A, denoted $\bar{A}$ or $\text{cl}(A)$, is defined as

$$\bar{A} := \bigcap_{A \subset F,\, F \text{ closed}} F\,.$$

Remark. We have $A^\circ \subset A \subset \bar{A}$. A° is the largest open subset of A and $\bar{A}$ is the smallest closed superset of A. $A^\circ = A$ if and only if A is open, and $\bar{A} = A$ if and only if A is closed.

The definition of closure and interior make the above facts very clear. However, while this clever definition makes it easy to prove the above properties, it is rather difficult to directly use it to compute the closure.

To facilitate this, we can define the limit points.

Definition 4.14. Let X be a topological space, $A \subset X$ and $x \in X$. x is called a *limit point* (*accumulation point*/*cluster point*) of A if for any neighbourhood U of x, $(A \setminus \{x\}) \cap U \neq \emptyset$.

The *derived set* of A is the set of all limit points of A, denoted A'.

Example. In $\mathbb{R}$ under the topology induced by the usual metric, suppose $A = [0, 1) \cup \{2\}$, then $A' = [0, 1]$. We also have $\mathbb{Q}' = \mathbb{R}$ and $\mathbb{Z}' = \emptyset$.

Proposition 4.15. Let (X, τ) be a topological space and $A \subset X$. Then A is closed if and only if $A' \subset A$.

Proof. If A is closed, then $U = X \setminus A$ is open, so for any $x \in X \setminus A$, U is a neighbourhood of x. But then $U \cap A = \emptyset$. Therefore, every limit point of A is inside of A, $A' \subset A$.

Conversely, given any $x \in X \setminus A$, we must have $x \notin A'$, so there is a neighbourhood U of x with $U \cap A = \emptyset$. Hence $x \in U \subset X \setminus A$, so $X \setminus A$ is open, and A is closed. □

Proposition 4.16. Let X be a topological space and $A \subset X$.

(i)
$$A^\circ = \{x \in X \mid A \text{ is a neighbourhood of } x\}\,;$$

(ii)
$$\begin{aligned}\bar{A} &= \{x \in X \mid \forall U \subset X \text{ such that } U \text{ is a neighbourhood of } x,\, U \cap A \neq \emptyset\} \\ &= A \cup A'\,.\end{aligned}$$

Proof.

(i) If A is a neighbourhood of x, then by definition there exists an open set U such that $x \in U \subset A$, which is true if and only if $x \in A^\circ$.

(ii) Suppose $x \notin \bar{A}$, then there exists a closed set $F \supset A$ such that $x \notin F$. Let $U = X \setminus F$. Then U is open and $x \in U$, so U is a neighbourhood of x and $U \cap A = \emptyset$.

Conversely, suppose there exists a neighbourhood U of x such that $U \cap A = \emptyset$. Then there exists an open set V such that $x \in V \subset U$. Since $V \subset U$, $V \cap A = \emptyset$. Let $F = X \setminus V$. Then F is closed, and $A \subset F$. Hence $\bar{A} \subset F$, so $x \notin \bar{A}$.

□

Example. In $\mathbb{R}$, let $A = [0,1) \cup \{2\}$. Then $A^\circ = (0,1)$ and $\bar{A} = [0,1] \cup \{2\}$. Further, $\mathbb{Q}^\circ = \emptyset$, $\bar{\mathbb{Q}} = \mathbb{R}$ and $\mathbb{Z}^\circ = \emptyset$, $\bar{\mathbb{Z}} = \mathbb{Z}$.

Remark. In a metric space, for a subset A we have that $x \in \bar{A}$ if and only if there exists a sequence x_n in A such that $x_n \to x$. In a general topological space, the existence of a sequence implies $x \in \bar{A}$, but the converse is not true.

Definition 4.17. A subset A of a topological space X is said to be *dense* in X if $\bar{A} = X$. X is *separable* if there exists a countable subset $A \subset X$ such that A is dense in X.

Example. $\mathbb{R}$ is separable as $\mathbb{Q}$ is dense in $\mathbb{R}$. $\mathbb{R}^n$ is separable as $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

Example. An uncountable discrete topological space is not separable, since the closure of any set is itself.

Definition 4.18. Let (X, τ) be a topological space. Let $Y \subset X$. Then the *subspace topology*, or *relative topology* on Y induced by τ is the topology
$$\{Y \cap V \mid V \in \tau\}\,,$$
the intersection of Y with all open sets in X. This is denoted by $\tau|_Y$.

For $U \subset Y$, U is open in Y if and only if there exists an open set $V \subset X$ such that $U = V \cap Y$.

Example. Let $X = \mathbb{R}$, $Y = [0,2]$ and $U = (1,2]$. Then $U \subset Y \subset X$. U is open in Y, since $V = (1,3)$ is open in X and $U = V \cap Y$. However, U is not open in X, since no neighbourhood around 2 can be constructed in X that is contained within U.

Remark. On a subset of a topological space, the subset topology is considered as the standard topology. Suppose that (X, τ) is a topological space, and $Z \subset Y \subset X$. There are two natural topologies on Z: $\tau|_Z$ and $\tau|_Y|_Z$. It can be easily shown that these two are equal.

Let (M, d) be a metric space and $N \subset M$. Again, there are two natural topologies on N: $\tau(d)|_N$ and $\tau(d|_N)$. These two constructions coincide. For any $x \in N$, $r > 0$,
$$\{y \in N \mid d(x,y) < r\} = \{y \in M \mid d(x,y) < r\} \cap N\,.$$

Proposition 4.19. Let X be a topological space, and let $A \subset Y \subset X$. A is closed in Y if and only if there is a closed subset $B \subset X$ such that $A = B \cap Y$. Further,

$$\mathrm{cl}_Y(A) = \mathrm{cl}_X(A) \cap Y\,.$$

Proof. The first part is true by taking complements. Let A be closed in Y, then $Y \setminus A$ is open in Y. Then $Y \setminus A = V \cap Y$ for some open V in X. So $B = X \setminus V$ is closed in X and $A = B \cap Y$. Conversely, if $A = B \cap Y$, where B is closed in X, then $X \setminus B$ is open in X, so $Y \setminus A = (X \setminus B) \cap Y$ is open in Y.

For the second part, we know $\mathrm{cl}_X(A)$ is closed in X, so by the first part, $\mathrm{cl}_X(A) \cap Y$ is closed in Y. Then $A \subset \mathrm{cl}_X(A) \cap Y$. So $\mathrm{cl}_Y(A) \subset \mathrm{cl}_X(A) \cap Y$. Similarly, since $\mathrm{cl}_Y(A)$ is closed in Y, we can write $\mathrm{cl}_Y(A) = B \cap Y$ for some closed set B in X. But $A \subset B$, and B is closed in X, so $\mathrm{cl}_X(A) \subset B$ and hence $\mathrm{cl}_Y(A) = B \cap Y \supset \mathrm{cl}_X(A) \cap Y$.

Remark. This is not true for the interior of a subset in general. For instance, consider $X = \mathbb{R}$, $A = Y = \{0\}$. Then $\mathrm{int}_Y(A) = A$ and $\mathrm{int}_X(A) = \emptyset$.

Remark. If $U \subset Y \subset X$, and Y is open in X, then U is open in Y if and only if U is open in X.

4.4 Continuity

4.4.1 Base

Definition 4.20. A *base* for a topological space (X, τ) is a family $\mathscr{B} \subset \tau$ such that $\forall U \in \tau$, $\exists \mathscr{C} \subset \mathscr{B}$ such that

$$U = \bigcup_{B \in \mathscr{C}} B\,.$$

In other words, the topology τ consists of the arbitrary unions of some family of open sets which is a subset of $\mathscr{B}$. So a base determines the topology.

Examples.

(i) The set of all open intervals is a base of the usual topology on $\mathbb{R}$.

(ii) In general, the collection of all open balls in a metric space is a base for the metric topology on it.

However, what we want to do is not to construct $\mathscr{B}$ from τ, but the other way around.

Lemma 4.21. Let X be a set and $\mathscr{B} \subset \mathcal{P}(X)$. Assume that

(i) $X = \bigcup_{B \in \mathscr{B}} B$;

(ii) $\forall B_1, B_2 \in \mathscr{B}$, $\forall x \in B_1 \cap B_2$, $\exists B \in \mathscr{B}$, $x \in B \subset B_1 \cap B_2$.

Then there is a unique topology on X that is generated by the base $\mathscr{B}$.

Proof. We must have the topology

$$\tau = \left\{ \bigcup_{B \in \mathscr{C}} B \;\middle|\; \mathscr{C} \subset \mathscr{B} \right\}.$$

It is clear that τ is a topology on X. Indeed, $\emptyset, X \in \tau$ and it is closed under arbitrary union. For intersection, consider

$$U_1 = \bigcup_{B \in \mathscr{C}_1} B\,,\; U_2 = \bigcup_{B \in \mathscr{C}_2} B\,.$$

Given that $x \in U_1 \cap U_2$, then $\exists B_1 \in \mathscr{C}_1$, $B_2 \in \mathscr{C}_2$ and $B_x \in \mathscr{B}$ such that $x \in B_x \subset B_1 \cap B_2 \subset U_1 \cap U_2$, thus

$$U_1 \cap U_2 = \bigcup_{x \in U_1 \cap U_2} B_x \,.$$

Definition 4.22. A topological space is called *second-countable* if it has a countable base.

Example. The set of all open balls of rational radii and centres is a countable base for $\mathbb{R}^n$, so $\mathbb{R}^n$ is second-countable.

4.4.2 Continuity

Definition 4.23. A function $f : X \to Y$ between topological spaces is *continuous* if for all open sets V in Y, the preimage $f^{-1}(V)$ is open in X.

Remark. We have already proven that this agrees with the definition of continuity of functions between metric spaces.

Examples.

(i) Constant functions are always continuous.

Consider $f : X \to Y$ defined by $f(x) = y_0$ for a fixed $y_0 \in Y$. For any $V \subset Y$, $f^{-1}(V) = \emptyset$ if $y_0 \notin V$, and $f^{-1}(V) = X$ if $y_0 \in V$. So f is continuous.

(ii) The identity map is always continuous.

If $f : X \to X$ is defined by $x \mapsto x$, $f^{-1}(V) = V$ so if V is open, then $f^{-1}(V)$ is trivially open.

(iii) Let $Y \subset X$. Let $\imath : Y \to X$ be the inclusion map. Then for an open set V in X, $\imath^{-1}(V) = V \cap Y$ which by definition is open in Y.

Proposition 4.24. Let $f : X \to Y$ be a function between topological spaces. Then

(i) f is continuous if for all closed sets B in Y, $f^{-1}(B)$ is closed in X;

(ii) if $\mathscr{B}$ is a base for Y, then f is continuous if and only if for all $B \in \mathscr{B}$, $f^{-1}(B)$ is open in X.

(iii) if f is continuous and $g : Y \to Z$ is continuous, then $g \circ f$ is also continuous.

Proof. Trivial. □

Remark. There exists a notion of *continuity at a point* for topological spaces, but it is not as important.

4.4.3 Homeomorphisms

Definition 4.25. A function $f : X \to Y$ between topological spaces is a *homeomorphism* if f is a bijection, and both f, f^{-1} are continuous.

If such an f exists, we say that X and Y are *homeomorphic*, written $X \cong Y$.

Definition 4.26. A function $f : X \to Y$ between topological spaces is an *open map* if for every U open in X, $f(U)$ is open in Y.

Remark. $f : X \to Y$ is a homeomorphism if and only if it is a continuous open bijection.

Lemma 4.27. Homeomorphism is an equivalence relation.

Proof.

(i) The identity map $I_X : X \to X$ is always a homeomorphism, so $X \cong X$.

(ii) If $f : X \to Y$ is a homeomorphism, then so is $f^{-1} : Y \to X$. So $X \cong Y \implies Y \cong X$.

(iii) If $f : X \to Y$ and $g : Y \to Z$ are homeomorphisms, then $g \circ f : X \to Z$ is a homeomorphism. So $X \cong Y$ and $Y \cong Z \implies X \cong Z$.

□

Examples.

(i) Under the usual topology, the open intervals $(0, 1) \cong (a, b)$ for all $a, b \in \mathbb{R}$ using the homeomorphism $x \mapsto a + (b - a)x$.

Similarly, $[0, 1] \cong [a, b]$.

(ii) $(-1, 1) \cong \mathbb{R}$ by $x \mapsto \tan\left(\frac{\pi}{2}x\right)$.

(iii) $\mathbb{R} \cong (0, \infty)$ by $x \mapsto \mathrm{e}^x$.

(iv) $(a, \infty) \cong (b, \infty)$ by $x \mapsto x + (b - a)$.

The fact that homeomorphism is an equivalence relation implies that any two open intervals in $\mathbb{R}$ are homeomorphic.

Remark. It is relatively easy to show that two spaces are homeomorphic. We just have to write down a homeomorphism. However, it is rather difficult to prove that two spaces are not homeomorphic.

For example, is $\mathbb{R}^n$ homeomorphic to $\mathbb{R}^m$ for $m \neq n$? The answer is they are not homeomorphic, but we are not able to prove it rigorously without using algebraic topology.

So how can we prove that two spaces are not homeomorphic? In group theory, we could prove that two groups are not isomorphic by, say, showing that they have different orders. Similarly, to distinguish between topological spaces, we have to define certain topological properties. Then if two spaces have different topological properties, we can show that they are not homeomorphic.

Definition 4.28. A property P of topological spaces is said to be a *topological property* or *topological invariant* if it is preserved under homeomorphisms.

Example. Metrizability is a topological invariant. Being Hausdorff is a topological invariant. Being completely metrizable (metrizable into a complete metric space) is a topological invariant. However completeness is not a topological invariant.

4.5 Product Topology and Quotient Topology

4.5.1 Product Topology

Let X, Y be topological spaces. We want to define the topology on $X \times Y$. If $U \subset X$ and $V \subset Y$ are open, then we want $U \times V$ to be open in $X \times Y$. Certainly, $\emptyset \times \emptyset = \emptyset$ and $X \times Y$ should be open. Further, $(U \times V) \cap (U' \times V') = (U \cap U') \times (V \cap V')$, so intersections work. $\bigcup_{i \in I} U_i \times V_i$ must be open for open sets U_i, V_i, but this is not obvious from what we have shown so far, so we must include this in our definition.

Or equivalently, we can define everything more neatly in terms of bases.

Definition 4.29. Let (X, τ) and (Y, ρ) be topological spaces. The base

$$\mathscr{B} = \{U \times V \mid U \in \tau, V \in \rho\}$$

defines a topology on $X \times Y$ called the *product topology*.

Remark. A set $W \subset X \times Y$ is open if and only if $\forall (x,y) \in W$, $\exists U \in \tau, V \in \rho$ such that $(x,y) \in U \times V \subset W$.

Example. Let (M,d), (M',d') be metric spaces. Then the d_∞ metric on $M \times M'$ is

$$d_\infty((x,x'),(y,y')) = \max\{d(x,y), d'(x',y')\}\,.$$

All d_p metrics induce the same metric topology on $M \times M'$, but this metric is chosen since it is easier to work with. Also, M and M' are topological spaces with their own metric topologies, which induce the product topology on the product space $M \times M'$. These two constructions create the same topology. For a point $z = (x,x') \in M \times M'$ and $r > 0$, the ball $D_r(z)$ is exactly

$$\begin{aligned} D_r(z) &= \{(y,y') \in M \times M' \mid d_\infty((y,y'),(x,x')) < r\} \\ &= \{(y,y') \in M \times M' \mid d(x,y) < r, d(x',y') < r\} \\ &= D_r(x) \times D_r(x')\,. \end{aligned}$$

Now let $W \subset M \times M'$. Then W is open in the product topology if and only if for all $z = (x,x') \in W$, there exists open sets $U \subset M$ and $U' \subset M'$ such that $(x,x') \in U \times U' \subset W$. Equivalently, for all $z = (x,x') \in W$, there exists $r > 0$ such that $D_r(x) \times D_r(x') \subset W$. But $D_r(x) \times D_r(x') = D_r(z)$, so W is d_∞-open as required.

Therefore, the topology induced by the (p-)product metric is the product topology of metric topologies. So products of metrizable topologies are metrizable.

For instance, $\mathbb{R}^2$ in the usual topology is homeomorphic to $\mathbb{R} \times \mathbb{R}$ in the product topology.

Proposition 4.30. Let X, Y be topological spaces. Let $X \times Y$ be given the product topology. Then the coordinate projections $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ satisfy

(i) π_X and π_Y are continuous;

(ii) if Z is any topological space, and $f : Z \to X \times Y$ is a function, then f is continuous if and only if $\pi_X \circ f$ and $\pi_Y \circ f$ are continuous.

Note that $f(z) = (\pi_X \circ f(z), \pi_Y \circ f(z))$.

Proof.

(i) Given an open set $U \subset X$, $\pi_X^{-1}(U) = U \times Y$, which is open in $X \times Y$ so π_X is continuous. Similarly, π_Y is continuous.

(ii) Given such an f, if f is continuous, then both $\pi_X \circ f$, $\pi_Y \circ f$ are continuous since composition of continuous functions is continuous.

Conversely, if both of $\pi_X \circ f$, $\pi_Y \circ f$ are continuous, then it is enough to check that any member of the base $U \times V \subset X \times Y$ has an open preimage. Indeed,

$$\begin{aligned} f^{-1}(U \times V) &= f^{-1}(U \times Y) \cap f^{-1}(X \times V) \\ &= f^{-1}(\pi_X^{-1}(U)) \cap f^{-1}(\pi_Y^{-1}(V)) \\ &= (\pi_X \circ f)^{-1}(U) \cap (\pi_Y \circ f)^{-1}(V) \end{aligned}$$

is open by assumption. □

Remark. The product topology may be extended to a finite product. Properties of the product topology hold. For example, if X_i is metrizable with metric d_i for all i, then the product topology is metrizable with, for instance, the d_∞ metric.

It is also useful to know that $(X \times Y) \times Z \cong X \times (Y \times Z)$, and $X \times Y \cong Y \times X$.

4.5.2 Quotient Topology

Definition 4.31. Let (X, τ) be a topological space and let R be an equivalence relation on X. Let X/R be the set of *equivalence classes*. Define the *quotient map*

$$\begin{aligned} q : X &\to X/R \\ x &\mapsto [x]\,, \end{aligned}$$

where $[x] \coloneqq \{y \in X \mid yRx\}$ is the equivalence class containing x. The *quotient topology* on X/R is the family

$$\tau_R = \{V \subset X/R \mid q^{-1}(V) \in \tau\}\,.$$

Proposition 4.32. The quotient topology is indeed a topology.

Proof. $q^{-1}(X/R) = X \in \tau$, $q^{-1}(\emptyset) = \emptyset \in \tau$, so $\emptyset, X/R \in \tau_R$. If $\forall i \in I$, $V_i \in \tau_R$, then

$$q^{-1}\left(\bigcup_{i\in I} V_i\right) = \bigcup_{i\in I} q^{-1}(V_i)\,,$$

which is open. Also, $\forall U, V \in \tau_R$,

$$q^{-1}(U \cap V) = q^{-1}(U) \cap q^{-1}(V)\,,$$

which is also open. □

Remarks.

(i) q is surjective and continuous under τ_R.

(ii) For $x \in X$, $t \in X/R$, $x \in t \iff q(x) = t$. Hence for $V \subset X/R$,

$$\begin{aligned} q^{-1}(V) &= \{x \in X \mid q(x) \in V\} \\ &= \{x \in X \mid \exists t \in V, q(x) = t\} \\ &= \bigcup_{t\in V} t\,. \end{aligned}$$

Example. Consider $\mathbb{R}$ as an abelian group under addition, and its subgroup $\mathbb{Z}$. We can form the quotient group $\mathbb{R}/\mathbb{Z}$, which is the set of equivalence classes where $x \sim y \iff x - y \in \mathbb{Z}$. For all $x \in \mathbb{R}$, there exists $y \in [0, 1]$ such that $x \sim y$, and for all $x, y \in [0, 1]$, we have $x \sim y$ if and only if $x = y$ or $\{x, y\} = \{0, 1\}$. So we can think of the quotient topology of $\mathbb{R}/\mathbb{Z}$ as a circle. We can say that $\mathbb{R}/\mathbb{Z}$ is homeomorphic to $S^1 = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| = 1\}$.

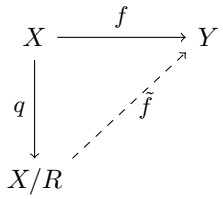
Example. Consider $\mathbb{Q} \leq \mathbb{R}$ as abelian groups under addition. $\mathbb{R}/\mathbb{Q}$ gives an equivalence relation. So we can induce a quotient topology on $\mathbb{R}/\mathbb{Q}$. Let $V \subset \mathbb{R}/\mathbb{Q}$, such that $V \neq \emptyset$ and V is open. Then $q^{-1}(V)$ is open and non-empty. Therefore, there exists $a < b \in \mathbb{R}$ such that $(a, b) \subset q^{-1}(V)$. Given $x \in \mathbb{R}$, we can choose a rational number r in the interval $(a - x, b - x)$. Then $r + x \in (a, b) \subset q^{-1}(V)$, so $q(x) = q(r + x) \in V$. So $V = \mathbb{R}/\mathbb{Q}$. We can find that this is the indiscrete topology, which is not metrizable or Hausdorff. So we cannot in general take quotients of metric spaces.

Example. Let $Q = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. We define the equivalence relationship by

$$(x_1, x_2) \sim (y_1, y_2) \iff \begin{cases} (x_1, x_2) = (y_1, y_2) & \text{or} \\ x_1 = y_1, \{x_2, y_2\} = \{0, 1\} & \text{or} \\ x_2 = y_2, \{x_1, y_1\} = \{0, 1\} & \text{or} \\ x_1, x_2, y_1, y_2 \in \{0, 1\} \end{cases}$$

The space Q/R is homeomorphic to $\mathbb{R}^2/\mathbb{Z}^2$. This is a square where the top and bottom edges are identified the same, and so do the left and right edges. This is homeomorphic to a surface of a torus with the Euclidean topology embedded in three-dimensional Euclidean space.

Consider the quotient map $q : X \to X/R$, and a map $f : X \to Y$ such that $xRy \implies f(x) = f(y)$, then there is a map $\tilde{f} : X/R \to Y$ such that $\tilde{f} \circ q = f$. In terms of category theory, the following diagram commutes.



If f is surjective, so is $\tilde{f}$. Also, if $f(x) = f(y) \iff xRy$, then $\tilde{f}$ is injective.

Proposition 4.33. Let X, Y be topological spaces, R an equivalence relation on X, $q : X \to X/R$ the quotient map, $f : X \to Y$ some map with $xRy \implies f(x) = f(y)$, and let $\tilde{f}$ be defined as above. Then

(i) if f is continuous, then so is $\tilde{f}$;

(ii) if f is an open map, then so is $\tilde{f}$.

Proof.

(i) Let V be open in Y, then $f^{-1}(V)$ is open in X, so $q^{-1}(\tilde{f}^{-1}(V)) = f^{-1}(V)$ is open, so $\tilde{f}^{-1}(V)$ is open, hence $\tilde{f}$ is continuous.

(ii) Given open $V \subset X/R$, $U = q^{-1}(V)$ is open in X, and $V = q(U)$, so $\tilde{f}(V) = f(U)$ which is open.

□

Corollary. If $f(x) = f(y) \iff xRy$, f is surjective, continuous and open, then $\tilde{f}$ is a homeomorphism.

Example. $\mathbb{R}/\mathbb{Z}$ is homeomorphic to a circle $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$.

Consider the map $f(t) = \mathrm{e}^{2\pi \mathrm{i} t}$, then the induced $\tilde{f}$ is a homeomorphism by the preceding corollary. If f is not open, then there is some open $U \in \mathbb{R}$ such that $f(U)$ is not open, and hence $S^1 \setminus f(U)$ is not closed. Then $\exists z_n \in S^1 \setminus f(U)$ and $z \in f(U)$ such that $z_n \to z$. Due to the surjectivity, we know that there is some $x \in U$ such that $f(x) = z$ and $x_n \in [x - 1/2, x + 1/2]$ such that $f(x_n) = z_n$. We know that $x_n \notin U$, but x_n have a convergent subsequence $x_{k_n} \to y \in \mathbb{R} \setminus U$ which is closed, so due to continuity, we must have $f(x) = f(y) \implies x - y \in \mathbb{Z} \implies x = y \notin U$, which is a contradiction.

Proposition 4.34. Let X be a topological space, and R an equivalence relation on X. Then

(i) if X/R is Hausdorff, then R is closed in $X \times X$;

(ii) if R is closed in $X \times X$ and the quotient map $q : X \to X/R$ is an open map, then X/R is Hausdorff.

Proof.

(i) Let $W = X \times X \setminus R$. We want to show W is open, so is a neighbourhood of all its points. Given $(x, y) \in W$, we have $x \not\sim y$, so $q(x) \neq q(y)$. Since the quotient is Hausdorff, there exists open sets S, T in X/R such that $S \cap T = \emptyset$ and $q(x) \in S$, $q(y) \in T$. Let $U = q^{-1}(S)$, $V = q^{-1}(T)$ which are open in X, and $x \in U$, $y \in V$. For all $(a, b) \in U \times V$, we have $q(a) \in S$, $q(b) \in T$, so $a \not\sim b$. So $(x, y) \in U \times V \subset W$. Hence R is closed.

(ii) Let $z \neq w \in X/R$, and we want to separate these points by open sets. Let $x, y \in X$ such that $q(x) = z$, $q(y) = w$. Then $(x, y) \in W$, which is open. Hence there exists open sets U, V in X such that $(x, y) \in U \times V \subset W$. Since q is an open map, $q(U)$ and $q(V)$ are open in X/R, and $z = q(x) \in q(U)$, $w = q(y) \in q(V)$. Now it suffices to show $q(U) \cap q(V) = \emptyset$, which is obvious as for all $(a, b) \in U \times V \subset W$, $(a, b) \notin R$, so $q(a) \neq q(b)$. □

5 Connectedness

5.1 Connectedness

An interval I in $\mathbb{R}$ has the defining property that $\forall x, y, z \in \mathbb{R}$, $x < y < z$, we have $x, z \in I \implies y \in I$. We know that a real continuous function maps intervals to intervals due to the intermediate value theorem. But it may not work if the (restricted) domain is not an interval.

Definition 5.1. A topological space is *disconnected* if there are open $U, V \subset X$ such that $U \neq \emptyset$ and $V \neq \emptyset$ partition X, i.e. $U \cap V = \emptyset$ and $U \cup V = X$. We then say U and V *disconnect* X.

A topological space X is *connected* if it is not disconnected.

Remark. Note that being connected is a property of a space, not a subset. When we say A is a connected subset of X, it means A is connected with the subspace topology inherited from X.

Being (dis)connected is a topological property, i.e. if X is (dis)connected, and $X \cong Y$, then Y is (dis)connected. To show this, let $f : X \to Y$ be the homeomorphism. By definition, A is open in X iff $f(A)$ is open in Y. So A and B disconnect X iff $f(A)$ and $f(B)$ disconnect Y.

Examples.

(i) If X has the indiscrete topology, it is connected.

(ii) If $|X| \geq 2$ and X has the discrete topology, then it is disconnected.

(iii) Let $X \subset \mathbb{R}$. If there is some $\alpha \in \mathbb{R} \setminus X$ such that there is some $a, b \in X$ with $a < \alpha < b$, then X is disconnected. In particular, $X \cap (-\infty, \alpha)$ and $X \cap (\alpha, \infty)$ disconnect X.

Proposition 5.2. X is disconnected $\iff$ there is a continuous surjective $f : X \to \{0, 1\}$ with the discrete topology.

Corollary. X is connected $\iff$ any continuous map $f : X \to \{0, 1\}$ is a constant.

Proof.

($\Rightarrow$) Let A and B disconnect X, and define

$$f(x) = \begin{cases} 0 & x \in A \\ 1 & x \in B \end{cases}.$$

$f^{-1}(\emptyset) = \emptyset$, $f^{-1}(\{0, 1\}) = X$, $f^{-1}(\{0\}) = A$ and $f^{-1}(\{1\}) = B$ are all open, so f is continuous.

($\Leftarrow$) Given such f surjective and continuous, define $A = f^{-1}(\{0\})$, $B = f^{-1}(\{1\})$. Then A and B disconnect X. □

Theorem 5.3. $[0, 1]$ is connected.

Note that $\mathbb{Q} \cap [0, 1]$ is disconnected. The key property of $\mathbb{R}$ is the least upper bound property.

Proof. Suppose A and B disconnect $[0, 1]$. Wlog, assume $1 \in B$. Since A is non-empty, $\alpha = \sup A$ exists. Then either

- $\alpha \in A$. Then $\alpha < 1$, since $1 \in B$. Since A is open, $\exists \epsilon > 0$ with $B\epsilon(\alpha) \subset A$. So $\alpha + \frac{\epsilon}{2} \in A$, contradicting the supremality of α; or
- $\alpha \notin A$. Then $\alpha \in B$. Since B is open, $\exists \epsilon > 0$ such that $B_\epsilon(\alpha) \subset B$. Then $a \leq \alpha - \epsilon$ for all $a \in A$. This contradicts α being the least upper bound of A.

So such A and B cannot exist and $[0,1]$ is connected.

Corollary. Let $X \subset \mathbb{R}$. X is connected if and only if it is an interval.

Lemma 5.4. If $f : X \to Y$ is continuous and X is connected. Then $\operatorname{im} f$ is also connected.

Proof. Let $f : X \to Y$ be continuous. Then if we alternatively consider f as $f : X \to \operatorname{im} f$, it is also continuous. Then if U, V disconnect $\operatorname{im} f$, $f^{-1}(U), f^{-1}(V)$ disconnect X.

Corollary. The following are equivalent:

(i) X is connected;

(ii) if $f : X \to \mathbb{R}$ is continuous, then $f(X)$ is an interval;

(iii) if $f : X \to \mathbb{Z}$ is continuous, then f is constant.

Theorem 5.5 (Intermediate value theorem). Suppose $f : X \to \mathbb{R}$ is continuous and X is connected. If $\exists x_0, x_1 \in X$ such that $f(x_0) < 0 < f(x_1)$, then $\exists x \in X$ with $f(x) = 0$.

Proof. Suppose no such x exists. Then $0 \notin \operatorname{im} f$ while $0 > f(x_0) \in \operatorname{im} f$, $0 < f(x_1) \in \operatorname{im} f$. Then $\operatorname{im} f$ is disconnected, contradicting X being connected. □

The converse of the intermediate value theorem is also true. If X is disconnected, we can always find a function g that does not satisfy the intermediate value property. Just let $f : X \to \{0, 1\}$ be continuous and define $g(x) = f(x) - \frac{1}{2}$.

Lemma 5.6. Let Y be a subspace of a topological space X. Then Y is disconnected if and only if there exists open subsets $U, V \subset X$ such that $U \cap V \cap Y = \emptyset$, $U \cup V \supset Y$, $U \cap Y \neq \emptyset$ and $V \cap Y \neq \emptyset$.

Proof. Suppose Y is disconnected. Then there exists open subsets $U', V' \subset Y$ that disconnect Y. Then there exists open sets $U, V \subset X$ such that $U' = U \cap Y$ and $V' = V \cap Y$. Then U, V satisfy the requirements.

Conversely, suppose U, V are as given. Then let $U' = U \cap Y$ and $V' = V \cap Y$. They disconnect Y. □

Proposition 5.7. Let Y be a subspace of a topological space X. If Y is connected, then so is $\bar{Y}$.

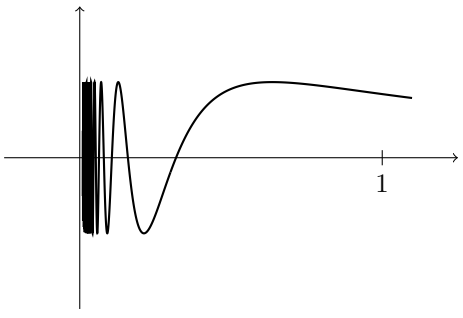
Proof. Suppose $\bar{Y}$ is disconnected. Then $\exists U, V \subset X$ open which disconnect $\bar{Y}$. Then $U \cap V \cap Y \subset U \cap V \cap \bar{Y} = \emptyset$. Hence $U \cap V \cap Y = \emptyset$. Also, $U \cup V \supset \bar{Y} \supset Y$. So by the previous lemma, U, V will disconnect Y unless $U \cap Y = \emptyset$ or $V \cap Y = \emptyset$. Since Y is connected, let us assume $V \cap Y = \emptyset$ wlog. Then $Y \subset X \setminus V$, which is closed, so we also have $\bar{Y} \subset X \setminus V$. Hence $V \cap \bar{Y} = \emptyset$. This is a contradiction since U, V disconnect $\bar{Y}$.

Proposition 5.8. Let X be connected and R an equivalence relation on X. Then the quotient space X/R is connected.

Proof. The quotient map is continuous and surjective. □

Example. Let

$$Y = \left\{ \left(x, \sin \frac{1}{x} \right) \,\middle|\, x > 0 \right\} \subset \mathbb{R}^2 .$$



This space is connected since it is the image of a continuous function

$$\begin{aligned} f : (0, \infty) &\to \mathbb{R}^2 \\ x &\mapsto \left(x, \sin \frac{1}{x}\right) \end{aligned}$$

on a connected space. Hence $\bar{Y}$ is connected. We claim that

$$Z \coloneqq Y \cup \{(0, y) \mid y \in [-1, 1]\}$$

is exactly $\bar{Y}$. Indeed, given $y \in [-1, 1]$, for all $n \in \mathbb{N}$ we have that $(0, 1/n)$ is mapped to (n, ∞) by $x \mapsto 1/x$, so by the intermediate value theorem there exists $x_n \in (0, 1/n)$ such that $\sin 1/x_n = y$. Hence,

$$\left(x_n, \sin \frac{1}{x_n}\right) = (x_n, y) \to (0, y) \in \bar{Y}\,,$$

so $Y \subset Z \subset \bar{Y}$. As Z is closed, it is exactly $\bar{Y}$.

Lemma 5.9. Let $\mathscr{A}$ be a family of connected subsets of a topological space X such that $\forall A, B \in \mathscr{A}$, $A \cap B \neq \emptyset$. Then

$$\bigcup_{A \in \mathscr{A}} A$$

is connected.

Proof. Let $Y = \bigcup_{A \in \mathscr{A}} A$ and let $f : Y \to \mathbb{Z}$ be continuous. For all $A \in \mathscr{A}$, $f|_A : A \to \mathbb{Z}$ is continuous and hence constant. For all $A, B \in \mathscr{A}$, $A \cap B \neq \emptyset$ hence $f|_A$ and $f|_B$ are both constant and have the same value. So f must be constant, and hence Y is connected. □

Proposition 5.10. Let X and Y be connected topological spaces. Then $X \times Y$ is connected.

Proof. Let $x_0 \in X$. Then $\{x_0\} \times Y \cong Y$ is connected. Similarly, for all $y \in Y$, $X \times \{y\} \cong X$ is connected. $\{x_0\} \times Y \cap X \times \{y\} = \{(x_0, y)\} \neq \emptyset$. Hence, $A_y = \{x_0\} \times Y \cup X \times \{y\}$ is connected by the previous lemma. For all $y, z \in Y$, $A_y \cap A_z \supset \{x_0\} \times Y$, so $A_y \cap A_z \neq \emptyset$. Hence, $\bigcup_{y \in Y} A_y = X \times Y$ is connected. □

Example. $\mathbb{R}^n$ is connected.

5.2 Path-connectedness

The other notion of connectedness is path-connectedness. A space is path-connected if we can join any two points with a path.

Definition 5.11. Let X be a topological space, and $x_0, x_1 \in X$. Then a *path* from x_0 to x_1 is a continuous function $\gamma : [0, 1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$.

Definition 5.12. A topological space is *path-connected* if for all points $x_0, x_1 \in X$, there is a path from x_0 to x_1.

Examples.

(i) (a, b), $(a, b]$, $[a, b)$ and $\mathbb{R}$ are all path-connected.

(ii) $\mathbb{R}^n$ is path-connected (e.g. using $\gamma(t) = tx_1 + (1 - t)x_0$).

(iii) $\mathbb{R}^n \setminus \{0\}$ is path-connected for $n > 1$. (Paths are either line segments or bent line segments to get around the hole).

Path-connectedness is a stronger condition than connectedness.

Proposition 5.13. If X is path-connected, then X is connected.

Proof. Let X be path-connected, and let $f : X \to \{0, 1\}$ be a continuous function. We will show that f is constant.

Let $x, y \in X$. By path-connectedness, there is a map $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Composing with f gives a map $f \circ \gamma : [0, 1] \to \{0, 1\}$. Since $[0, 1]$ is connected, the map must be a constant. In particular, $f(\gamma(0)) = f(\gamma(1))$, i.e. $f(x) = f(y)$. Since x, y were arbitrary, f is constant.

The converse is not true.

Example. Take the topologist's sine wave,

$$X = \left\{ \left(x, \sin \frac{1}{x} \right) \,\middle|\, x > 0 \right\} \cup \{(0, y) \mid y \in [-1, 1]\}.$$

We have shown that it is connected, but it is not path-connected.

To show this, pick points $(0, 0), (1, \sin(1)) \in X$. Assume that $\gamma : [0, 1] \to X$ is continuous and $\gamma(0) = (0, 0) = x$, $\gamma(1) = (1, \sin 1) = y$. Let $\gamma = (\gamma_1, \gamma_2)$, so γ_1 and γ_2 are continuous. For $t \in [0, 1]$ such that $\gamma_1(t) > 0$, $[0, \gamma_1(t)] \subset \gamma_1([0, t])$ by the intermediate value theorem, so $\exists n \in \mathbb{N}$ such that $(2\pi n)^{-1}, (2\pi n + \pi/2)^{-1} \in (0, \gamma_1(t)) \subset \gamma_1([0, t])$. So there is some $a, b < t$ with $\gamma_1(a) = (2\pi n)^{-1}$, $\gamma_1(b) = (2\pi n + \pi/2)^{-1}$, hence $\gamma_2(a) = 0$, $\gamma_2(b) = 1$, so we can thus find a sequence $1 > t_1 > t_2 > \cdots > 0$ with

$$\gamma_2(t_n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

So t_n converges but $\gamma_2(t_n)$ does not. This is a contradiction.

We can use connectivity to distinguish spaces. Apart from the obvious "X is connected while Y is not", we can also try to remove points and see what happens.

Lemma 5.14. Suppose $f : X \to Y$ is a homeomorphism and $A \subset X$, then $f|_A : A \to f(A)$ is a homeomorphism.

Proof. Since f is a bijection, $f|_A$ is a bijection. If $U \subset f(A)$ is open, then $U = f(A) \cap U'$ for some U' open in Y. So $f|_A^{-1}(U) = f^{-1}(U) \cap A$ is open in A. So $f|_A$ is continuous. Similarly, we can show that $f|_A^{-1}$ is continuous so $f|_A$ is a homeomorphism. □

Example. $[0, 1] \ncong (0, 1)$.

Suppose it were. Let $f : [0, 1] \to (0, 1)$ be a homeomorphism. Let $A = (0, 1]$. Then $f|_A : (0, 1] \to (0, 1) \setminus \{f(0)\}$ is a homeomorphism. But $(0, 1]$ is connected while $(0, 1) \setminus \{f(0)\}$ is disconnected. Contradiction.

Similarly, $[0, 1) \ncong [0, 1]$ and $[0, 1) \ncong (0, 1)$. Also, $\mathbb{R}^n \ncong \mathbb{R}$ for $n > 1$, and S^1 is not homeomorphic to any subset of $\mathbb{R}$.

Higher Connectivity*

We were able to use path-connectedness to determine that $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^n$ for $n > 1$. If we want to distinguish general $\mathbb{R}^n$ and $\mathbb{R}^m$, we will need to generalise the idea of path-connectedness to higher connectivity.

We have to formulate path-connectedness in a different way. Recall $S^0 = \{-1, 1\} \cong \{0, 1\} \subset \mathbb{R}$ and $D^1 = [-1, 1] \cong [0, 1] \subset \mathbb{R}$. Then we can formulate path-connectedness as: X is path-connected if any continuous $f : S^0 \to X$ extends to a continuous $\gamma : D^1 \to X$ with $\gamma|_{S^0} = f$.

Definition 5.15. X is *n-connected* if for any $k \leq n$, any continuous $S^k \to X$ extends to a continuous $F : D^{k+1} \to X$ such that $F|_{S^k} = f$.

For any point $p \in \mathbb{R}^n$, $\mathbb{R}^n \setminus \{p\}$ is m-connected if and only if $m \leq n-2$. So $\mathbb{R}^n \setminus \{p\} \ncong \mathbb{R}^m \setminus \{q\}$ unless $n = m$. So $\mathbb{R}^n \ncong \mathbb{R}^m$.

5.3 Components

If a space is disconnected, we could divide the space into different components, each of which is (maximally) connected. We would have a different notion of components for each type of connectivity.

5.3.1 Path Components

Lemma 5.16 (Gluing lemma). Let $f : X \to Y$ be a function between topological spaces. If $X = A \cup B$ where A, B are closed and $f|_A, f|_B$ are continuous, then f is continuous.

Proof. Given V closed in Y,

$$f^{-1}(V) = (f^{-1}(V) \cap A) \cup (f^{-1}(V) \cap B) = f|_A^{-1}(V) \cup f|_B^{-1}(V)$$

is closed since A, B are closed. Hence f is continuous. □

Lemma 5.17. Define $x \sim y$ if there is a path from x to y in X. $\sim$ is an equivalence relation.

Proof.

(i) For any $x \in X$, $x \sim x$ using the constant path.

(ii) If $\gamma : [0,1] \to X$ is a path from x to y, then $\bar{\gamma} : [0,1] \to X$ by $t \to \gamma(1-t)$ is a path from y to x. So $x \sim y \implies y \sim x$.

(iii) If γ_1 is a path from x to y and γ_2 is a path from $y \to z$. Then $\gamma_2 * \gamma_1$ defined by

$$t \mapsto \begin{cases} \gamma_1(2t) & t \in [0, 1/2] \\ \gamma_2(2t-1) & t \in [1/2, 1] \end{cases}$$

is a path from x to z. So $x \sim y, y \sim z \implies x \sim z$.

□

We can then immediately define the path components.

Definition 5.18. Equivalence classes of the relation "$x \sim y$ if there is a path from x to y" are *path components* of X.

5.3.2 Connected Components

Definition 5.19. For $x \in X$, define

$$\mathcal{C}(x) = \{A \subset X \mid x \in A \text{ and } A \text{ is connected}\}.$$

Then $C_x = \bigcup_{A \in \mathcal{C}(x)} A$ is the *connected component* of x.

C_x is the largest connected subset of X containing x. As $\{x\} \in \mathcal{C}(x)$, $x \in C_x$. Note that $\{x\} \in \bigcap_{A \in \mathcal{C}(x)} A$, so it is non-empty. As each $A \in \mathcal{C}(x)$ is connected, C_x is connected by a lemma above.

Lemma 5.20. If $y \in C_x$, then $C_y = C_x$.

Proof. Since $y \in C_x$ and C_x is connected, $C_x \subset C_y$. So $x \in C_y$. Then $C_y \subset C_x$. Hence $C_x = C_y$. □

It follows that $x \sim y$ if $x \in C_y$ defines an equivalence relation and the connected components of X are the equivalence classes.

Examples.

(i) Let $X = (-\infty, 0) \cup (0, \infty) \subset \mathbb{R}$. Then the connected components are $(-\infty, 0)$ and $(0, \infty)$, which are also the path components.

(ii) Let $X = \mathbb{Q} \subset \mathbb{R}$. Then $C_x = \{x\}$ for all $x \in X$. In this case, we say X is *totally disconnected*.

It is also important to note that path components need not be equal to the connected components, as illustrated by the following example. However, since path-connected spaces are connected, the path component containing x must be a subset of C_x.

The topologist's sine function would be an example. The whole function is connected, but it divides into two path-connected components.

Finally, recall that we showed that path-connected subsets are connected. While the converse is not true in general, there are special cases where it is true.

Proposition 5.21. If $U \subset \mathbb{R}^n$ is open and connected, then it is path-connected.

Proof. Let A be a path component of U. We first show that A is open.

Let $a \in A$. Since U is open, $\exists \epsilon > 0$ such that $D_\epsilon(a) \subset U$. We know that $D_\epsilon(a)$ is path-connected. Since A is a path component and $a \in A$, we must have $D_\epsilon(a) \subset A$. So A is an open subset of U.

Next suppose $b \in U \setminus A$. Then since U is open, $\exists \epsilon > 0$ such that $D_\epsilon(b) \subset U$. Since $D_\epsilon(b)$ is path-connected, so if $D_\epsilon(b) \cap A \neq \emptyset$, then $D_\epsilon(b) \subset A$. But this implies $b \in A$, which is a contradiction. So $D_\epsilon(b) \cap A = \emptyset$. So $D_\epsilon(b) \subset U \setminus A$. Then $U \setminus A$ is open.

So A, $U \setminus A$ are open subsets of U. Since U is connected, we must have $U \setminus A$ empty, so $U = A$ is path-connected. □

6 Compactness

Compactness is an important concept in topology. It can be viewed as a generalisation of being closed and bounded in $\mathbb{R}$. Alternatively, it can also be viewed as a generalisation of being finite. However, there are no corresponding definitions in general metric spaces or vector spaces.

Compact sets tend to have a lot of really nice properties.

Examples.

(i) If X is finite, then any function $X \to \mathbb{R}$ is bounded.

(ii) For a $f : X \to \mathbb{R}$, if there exists $n \in \mathbb{N}$ and $A_1, \dots A_n \subset X$ such that $X = \bigcup_i A_i$ and f is bounded on each A_i, then f is bounded on X.

(iii) Continuous functions are 'locally bounded'. If $f : X \to \mathbb{R}$ is continuous, then for all $x \in X$, we have $U_x = f^{-1}((f(x)-1, f(x)+1))$ an open set containing x, and f is bounded on U_x. So each point has an open neighbourhood on which f is bounded. Further, $X = \bigcup_{x\in X} U_x$. If there exists a finite subset $F \subset X$ such that $\bigcup_{x\in F} U_x = X$, then f is bounded on X.

This is exactly what we will use to define compactness.

6.1 Compactness

Definition 6.1. Let (X, τ) be a topological space. An *open cover* of X is a subset $\mathscr{U} \subset \tau$ such that

$$\bigcup_{U\in\mathscr{U}} U = X\,.$$

We say $\mathscr{U}$ covers X.

If $\mathscr{V} \subset \mathscr{U}$ and $\mathscr{V}$ covers X, then we say that $\mathscr{V}$ is a *subcover* of $\mathscr{U}$.

Definition 6.2. A topological space X is *compact* if every open cover $\mathscr{U}$ of X has a finite subcover $\mathscr{V} = \{V_1, \dots, V_n\} \subset \mathscr{U}$.

Examples.

(i) Let X be a finite set. Then any cover $\mathscr{U} \subset \tau \subset \mathcal{P}(X)$ is finite. So X is compact.

(ii) Let $X = \mathbb{R}$ and $\mathscr{U} = \{(-R, R) \mid R \in \mathbb{R}, R > 0\}$. This is an open cover with no finite subcover. So $\mathbb{R}$ is not compact. Hence all open intervals are not compact since they are homeomorphic to $\mathbb{R}$.

(iii) Let $X = [0, 1] \cap \mathbb{Q}$. Let
$$U_n = X \setminus (\alpha - 1/n, \alpha + 1/n)$$
for some irrational number $\alpha \in (0, 1)$. Then $\bigcup_{n>0} U_n = X$ since α is irrational. Then $\mathscr{U} = \{U_n \mid n \in \mathbb{Z}, n > 0\}$ is an open cover of X. Since it has no finite subcover, X is not compact.

Theorem 6.3. $[0, 1]$ is compact.

Again, this is not true for $[0, 1] \cap \mathbb{Q}$ so we must use some special properties of the reals.

Proof. Suppose $\mathscr{U}$ is an open cover of $[0, 1]$. Let

$$A = \{a \in [0, 1] \mid [0, a] \text{ has a finite subcover of } \mathscr{U}\}\,.$$

First we can show that A is non-empty. Since $\mathscr{U}$ covers $[0, 1]$, there must be some $U_0 \in \mathscr{U}$ that contains 0. So $\{0\}$ has a finite subcover $\mathscr{U}_0 = \{U_0\}$. So $0 \in A$.

Next we note that by definition, if $0 \leq b \leq a$ and $a \in A$, then $b \in A$. Now, let $\alpha = \sup A$. Then $\alpha \in [0, 1]$.

Since $\mathscr{U}$ covers $X = [0, 1]$, let $\alpha \in U_\alpha \in \mathscr{U}$. Since U_α is open, there is some ϵ such that $D_\epsilon(\alpha) \subset U_\alpha$. By the definition of α, we must have $\alpha - \epsilon/2 \in A$, so $[0, \alpha - \epsilon/2]$ has a finite subcover. Add U_α to that subcover to get a finite subcover of $[0, \eta]$, where $\eta = \min\{\alpha + \epsilon/2, 1\}$. This is a contradiction unless $\alpha = \sup A = 1$.

Also, $\exists U_1 \in \mathscr{U}$ such that $1 \in U_1$ and $\epsilon > 0$ such that $(1 - \epsilon, 1] \subset U_1$. Since $1 - \epsilon \in A$, there exists a finite $\mathscr{V} \subset \mathscr{U}$ that covers $[0, 1 - \epsilon/2]$. Then $\mathscr{W} = \mathscr{V} \cup \{U_1\} \subset \mathscr{U}$ is a finite subcover. □

We mentioned that compactness is a generalisation of "closed and bounded". We will now show that compactness is indeed in some ways related to closedness.

Proposition 6.4. If X is compact and C is a closed subset of X, then C is also compact.

Proof. Suppose $\mathscr{U}$ is an open cover of C. Let $\mathscr{U} = \{U_\alpha \mid \alpha \in T\}$. For each α, since U_α is open in C, $U_\alpha = C \cap U'_\alpha$ for some U'_α open in X. Also, since $\bigcup_{\alpha \in T} U_\alpha = C$, we have $\bigcup_{\alpha \in T} U'_\alpha \supset C$.

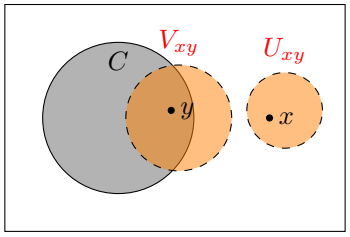
Since C is closed, $U = X \setminus C$ is open in X, so $\mathscr{V} = \{U'_\alpha \mid \alpha \in T\} \cup \{U\}$ is an open cover of X. Since X is compact, $\mathscr{V}$ has a finite subcover $\mathscr{V}' = \{U'_{\alpha_1}, \dots, U'_{\alpha_n}, U\}$ (U may or may not be there, but it does not matter). Since $U \cap C = \emptyset$, $\mathscr{U}' = \{U_{\alpha_1}, \dots, U_{\alpha_n}\} \subset \mathscr{U}$ is a finite subcover of C. □

The converse is not always true, but holds for Hausdorff spaces.

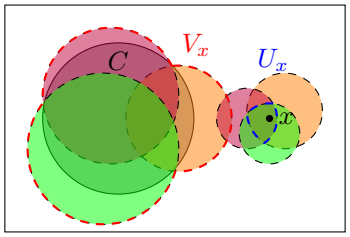
Proposition 6.5. Let X be a Hausdorff space. If $C \subset X$ is compact, then C is closed in X.

Proof. Let $U = X \setminus C$. We will show that U is open. For any $x \in U$, we will find a U_x such that $x \in U_x \subset U$. Then $U = \bigcup_{x \in U} U_x$ will be open since it is a union of open sets.

To construct U_x, fix $x \in U$. Since X is Hausdorff, for each $y \in C$, $\exists U_{xy}, V_{xy}$ as open neighbourhoods of x and y respectively with $U_{xy} \cap V_{xy} = \emptyset$.



Then $\mathscr{U} = \{V_{xy} \cap C \mid y \in C\}$ is an open cover of C. Since C is compact, there exists a finite subcover $\mathscr{V} = \{V_{xy_1} \cap C, \dots, V_{xy_n} \cap C\}$.



Let $U_x = \bigcap_{i=1}^n U_{xy_i}$. Then U_x is open since it is a finite intersection of open sets. To show $U_x \subset U$, note that $V_x = \bigcup_{i=1}^n V_{xy_i} \supset C$ since $\mathscr{V} = \{V_{xy_i} \cap C\}$ is an open cover. We also have $V_x \cap U_x = \emptyset$. So $U_x \subset U$. So done. □

After relating compactness to closedness, we will relate it to boundedness. First we need to define boundedness for general metric spaces.

Definition 6.6. A metric space (M, d) is *bounded* if there exists $L \in \mathbb{R}$ such that $d(x, y) \leq L$ for all $x, y \in M$.

Example. $A \subset \mathbb{R}$ is bounded iff $A \subset [-N, N]$ for some $N \in \mathbb{R}$.

Remark. Note that being bounded is not a topological property. For example, $(0, 1) \cong \mathbb{R}$ but $(0, 1)$ is bounded while $\mathbb{R}$ is not. It depends on the metric d, not just the topology it induces.

Proposition 6.7. A compact metric space is bounded.

Proof. Pick $x \in X$. Then $\mathscr{U} = \{D_r(x) \mid r \in \mathbb{R}^+\}$ is an open cover of X. Since X is compact, it has a finite subcover $\{D_{r_1}(x), \dots, D_{r_n}(x)\}$. Let $R = \max\{r_1, \dots, r_n\}$. Then $d(x, y) < R$ for all $y \in X$. So for all $y, z \in X$,

$$d(y, z) \leq d(x, y) + d(x, z) < 2R\,,$$

so X is bounded.

Theorem 6.8 (Heine–Borel theorem). $C \subset R$ is compact if and only if C is closed and bounded.

Proof. Since $\mathbb{R}$ is a metric space, it is Hausdorff and C is also a metric space.

If C is compact, then C is closed in $\mathbb{R}$ and is bounded by the previous two propositions.

Conversely, if C is closed and bounded, then $C \subset [-N, N]$ for some $N \in \mathbb{R}$. Since $[-N, N] \cong [0, 1]$ is compact, and C is closed in $[-N, N]$, C is compact. □

Corollary. If $A \subset \mathbb{R}$ is compact, $\exists \alpha \in A$ such that $\alpha \geq a$ for all $a \in A$.

Proof. Since A is compact, it is bounded. Let $\alpha = \sup A$. Then by definition, $\alpha \geq a$ for all $a \in A$. So it is enough to show that $\alpha \in A$.

Suppose $\alpha \notin A$. Then $\alpha \in \mathbb{R} \setminus A$. Since A is compact, it is closed in $\mathbb{R}$, so $\mathbb{R} \setminus A$ is open. So $\exists \epsilon > 0$ such that $D_\epsilon(\alpha) \subset \mathbb{R} \setminus A$, which implies that $a \leq \alpha - \epsilon$ for all $a \in A$. This contradicts the assumption that $\alpha = \sup A$. So we can conclude $\alpha \in A$. □

We call $\alpha = \max A$ the *maximum element* of A.

We have previously proved that if X is connected and $f : X \to Y$ is continuous, then $\operatorname{im} f \subset Y$ is connected. The same statement is true for compactness.

Proposition 6.9. If $f : X \to Y$ is continuous and X is compact, then $\operatorname{im} f \subset Y$ is also compact.

Proof. Suppose $\mathscr{U} = \{U_\alpha \mid \alpha \in T\}$ is an open cover of $\operatorname{im} f$. Since U_α is open in $\operatorname{im} f$, we have $V_\alpha = f^{-1}(U_\alpha)$ open in X. If $x \in X$ then $f(x)$ is in U_α for some α, so $x \in V_\alpha$. Thus $\mathscr{V} = \{V_\alpha \mid \alpha \in T\}$ is an open cover of X.

Since X is compact, there is a finite subcover $\mathscr{V}' = \{V_{\alpha_1}, \dots, V_{\alpha_n}\}$ of $\mathscr{V}$. Since $U_\alpha \subset \operatorname{im} f$, $f(V_\alpha) = f(f^{-1}(U_\alpha)) \subseteq U_\alpha$. So

$$\mathscr{U}' = \{U_{\alpha_1}, \dots, U_{\alpha_n}\}$$

is a finite subcover of $\mathscr{U}$. □

Theorem 6.10 (Maximum value theorem). If $f : X \to \mathbb{R}$ is continuous and X is compact, then $\exists x \in X$ such that $f(x) \geq f(y)$ for all $y \in X$.

Proof. Since X is compact, $\operatorname{im} f$ is compact. Let $\alpha = \max\{\operatorname{im} f\}$. Then $\alpha \in \operatorname{im} f$. So $\exists x \in X$ with $f(x) = \alpha$. Then by definition $f(x) \geq f(y)$ for all $y \in X$. □

Remark. A function on a compact domain is bounded and attains its bound.

6.2 Products and Quotients

6.2.1 Products

Recall the product topology on $X \times Y$: $U \subset X \times Y$ is open if it is a union of sets of the form $V \times W$ such that $V \subset X$, $W \subset Y$ are open.

The major takeaway of this section is the following theorem.

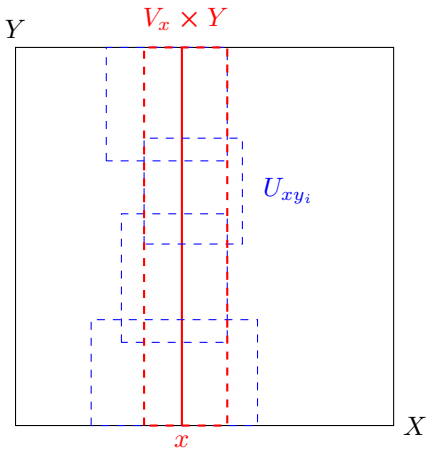
Theorem 6.11. If X and Y are compact, then so is $X \times Y$.

Proof. First consider the special type of open cover $\mathscr{U}$ of $X \times Y$ such that every $U \in \mathscr{U}$ has the form $U = V \times W$, where $V \subset X$ and $W \subset Y$ are open. For every $(x, y) \in X \times Y$, there is $U_{xy} \in \mathscr{U}$ with $(x, y) \in U_{xy}$. Write

$$U_{xy} = V_{xy} \times W_{xy}\,,$$

where $V_{xy} \subset X$, $W_{xy} \subset Y$ are open, $x \in V_{xy}$, $y \in W_{xy}$.

Fix $x \in X$. Then $\mathscr{W}_x = \{W_{xy} \mid y \in Y\}$ is an open cover of Y. Since Y is compact, there is a finite subcover $\{W_{xy_1}, \dots, W_{xy_n}\}$. Then $V_x = \bigcap_{i=1}^n V_{xy_i}$ is a finite intersection of open sets. So V_x is open in X. Moreover, $\mathscr{U}_x = \{U_{xy_1}, \dots, U_{xy_n}\}$ covers $V_x \times Y$.



Now $\mathscr{V} = \{V_x \mid x \in X\}$ is an open cover of X. Since X is compact, there is a finite subcover $\mathscr{V}' = \{V_{x_1}, \dots, V_{x_m}\}$. Then $\mathscr{U}' = \bigcup_{i=1}^m \mathscr{U}_{x_i}$ is a finite subset of $\mathscr{U}$, which covers all of $X \times Y$.

In the general case, suppose $\mathscr{U}$ is an open cover of $X \times Y$, For each $(x, y) \in X \times Y$, $\exists U_{xy} \in \mathscr{U}$ with $(x, y) \in U_{xy}$. Since U_{xy} is open, $\exists V_{xy} \subset X$, $W_{xy} \subset Y$ with $V_{xy} \times W_{xy} \subset U_{xy}$ and $x \in V_{xy}, y \in W_{xy}$.

Then $\mathscr{Q} = \{V_{xy} \times W_{xy} \mid (x, y) \in X \times Y\}$ is an open cover of $X \times Y$ of the type we already considered above. So it has a finite subcover $\{V_{x_1y_1} \times W_{x_1y_1}, \dots V_{x_my_n} \times W_{x_my_n}\}$. Now $V_{x_iy_i} \times W_{x_iy_i} \subset U_{x_iy_i}$. So $\{U_{x_iy_i}, \dots, U_{x_my_n}\}$ is a finite subcover of $\mathscr{U}$. □

Example. The unit cube $[0, 1]^n = [0, 1] \times \dots \times [0, 1]$ is compact.

Corollary (Heine–Borel theorem in $\mathbb{R}^n$). $C \subset \mathbb{R}^n$ is compact iff C is closed and bounded.

Proof. If C is bounded, $C \subset [-N, N]^n$ for some $N \in \mathbb{R}$, which is compact. The rest of the proof is the same as for $n = 1$. □

Corollary (Tychonoff's theorem (finite product case)). Finite products of compact spaces are compact.

Remark. This also works for arbitrary products, but the proof is much harder.

6.2.2 Quotients

Proposition 6.12. If X is compact and R is an equivalence relation on X, then X/R is compact.

Proof. The quotient map is continuous and surjective, so its image is compact. □

Theorem 6.13 (Topological inverse function theorem). Suppose $f : X \to Y$ is a continuous bijection. If X is compact and Y is Hausdorff, then f is a homeomorphism.

Proof. We show that f^{-1} is continuous. To do this, it suffices to show $(f^{-1})^{-1}(C)$ is closed in Y whenever C is closed in X. By hypothesis, f is a bijection, so $(f^{-1})^{-1}(C) = f(C)$. Suppose C is closed in X. Since X is compact, C is compact. Since f is continuous, $f(C) = \operatorname{im} f|_C$ is compact. Since Y is Hausdorff and $f(C) \subset Y$ is compact, $f(C)$ is closed as required. □

We will apply this to quotients.

Recall that if R is an equivalence relation on X, $q : X \to X/R$ is continuous if and only if $f : X \to Y$ is continuous, where $f \equiv \tilde{f} \circ q$ is defined such that $xRy \implies f(x) = f(y)$.

Corollary. Suppose $\tilde{f} : X/R \to Y$ is a bijection, X is compact, Y is Hausdorff, and $\tilde{f} \circ q$ is continuous, then $\tilde{f}$ is a homeomorphism.

Proof. Since X is compact and $q : X \to X/R$ is continuous, $\operatorname{im} q \subset X/R$ is compact. Since $\tilde{f} \circ q$ is continuous, $\tilde{f}$ is continuous, so we can apply the above proposition.

Example. Let $X = D^2$ and $A = S^1 \subset X$. Then

$$\begin{aligned} \tilde{f} : X/A &\to S^2 \\ (r, \theta) &\mapsto (1, \pi r, \theta) \end{aligned}$$

in spherical coordinates is a homeomorphism.

We can check that $\tilde{f}$ is a continuous bijection and D^2 is compact. So $D^2/S^1 \cong S^2$.

6.3 Sequential Compactness

The other definition of compactness is sequential compactness. We will not do much with it, but only prove that it is the same as compactness for metric spaces.

Definition 6.14. A topological space X is *sequentially compact* if every sequence $x_n \in X$ has a convergent subsequence.

Example. Any closed and bounded subset of $\mathbb{R}^n$ is sequentially compact by Bolzano–Weierstrass theorem.

Lemma 6.15. Let x_n be a sequence in a metric space (M, d) and $x \in M$. Then x_n has a subsequence converging to x if and only if for every $\epsilon > 0$, $x_n \in D_\epsilon(x)$ for infinitely many n.

Proof.

($\Rightarrow$) If $x_{n_i} \to x$, then for every $\epsilon > 0$, we can find $I \in \mathbb{N}$ such that $i > I$ implies $x_{n_i} \in D_\epsilon(x)$.

($\Leftarrow$) Now suppose for every $\epsilon > 0$, there are infinitely many $x_n \in D_\epsilon(x)$. We will construct a sequence x_{n_i} inductively. Take $n_0 = 0$. Suppose we have defined $x_{n_0}, \dots, x_{n_{i-1}}$. By hypothesis $x_n \in D_{1/i}(x)$ for infinitely many n. Take n_i to the smallest such n with $n_i > n_{i-1}$. Then $d(x_{n_i}, x) < 1/i$ implies that $x_{n_i} \to x$. □

Definition 6.16. Let (M, d) be a metric space. For $\epsilon > 0$ and $F \subset M$. We say F is an *ϵ-net* for M if $\forall x \in M$, $\exists y \in F$ such that $d(x, y) \leq \epsilon$. That is,

$$M = \bigcup_{y \in F} B_\epsilon(y)\,.$$

We say M is *totally bounded* if for any $\epsilon > 0$, there is a finite ϵ-net for M.

Example. For $\epsilon > 0$, take n such that $1/n < \epsilon$. Then

$$\left\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right\}$$

is an ϵ-net for $(0, 1)$.

Note that any compact space is totally bounded, but the converse is not true. The $(0, 1)$ above is an example. The only thing missing here is completeness.

Definition 6.17. For a non-empty set $A \subset M$, the *diameter* of A is

$$\operatorname{diam} A \coloneqq \sup_{x,y \in A} d(x, y)\,.$$

Clearly, the diameter of a set is finite if and only if it is bounded.

Theorem 6.18. For a metric space, the following are equivalent:

(i) M is compact.

(ii) M is sequentially compact.

(iii) M is totally bounded and complete.

Proof.

- (i) $\Rightarrow$ (ii). Let x_n be a sequence in M, so for $n \in \mathbb{N}$, let $A_n = \{x_k \mid k > N\}$. Note that the limit of any convergent subsequence, if it exists, is in the intersection $\bigcap_{n \in \mathbb{N}} \bar{A}_n$. We first show that it is non-empty. Assume it is empty, then

  $$\bigcup_{n \in \mathbb{N}} M \setminus \bar{A}_n = M\,.$$

  Each $M \setminus \bar{A}_n$ is open, so it is an open cover of M. By the compactness of M, it has a finite subcover, so there is some $N \in \mathbb{N}$ such that $\bigcup_{n \leq N} M \setminus \bar{A}_n = M$. But $\{A_n\}$ is decreasing, so necessarily $M \setminus \bar{A}_N = M$, then $\bar{A}_N = \emptyset$. Contradiction.

  Now let $x \in \bigcap_{n \in \mathbb{N}} \bar{A}_n$, and we want to show the existence of a subsequence converging to x. First, $x \in \bar{A}_1$, so $D_1(x) \cap A_1 \neq \emptyset$. Hence there exists $k_1 > 1$ such that $d(x_{k_1}, x) < 1$. Now since $x \in \bar{A}_{k_1}$, $D_{1/2}(x) \cap A_{k_1} \neq \emptyset$, there exists $k_2 > k_1$ such that $d(x_{k_2}, x) < 1/2$. Inductively, we can find $k_1 < k_2 < \ldots$ such that $d(x_{k_n}, x) < \frac{1}{n}$ for all $n \in \mathbb{N}$. Hence $x_{k_n} \to x$.

- (ii) $\Rightarrow$ (iii). M is complete since a Cauchy sequence with a convergent subsequence is convergent. To see it is totally bounded, assume it is not, then there is some $\epsilon > 0$ such that every ϵ-net is infinite. Pick any $x_1 \in M$, and once we have already picked $x_1, \ldots, x_n$, we pick $x_{n+1} \notin \bigcup_{k=1}^{n} B_\epsilon(x_k)$. This is always valid since otherwise M would have a finite ϵ-net. But then the sequence x_n cannot have any Cauchy subsequence, so it has no converging subsequence.

- (iii) $\Rightarrow$ (i). Assume such an M is not compact, so there is an open cover $\mathscr{U}$ without any finite subcover. We say $A \subset M$ is "bad" if there is no finite subcover of A in $\mathscr{U}$. So, for example, $M \subset M$ is bad while $\emptyset \subset M$ is not. Note if $A = \cup_{i=1}^n B_i$ is bad, then there is some i such that B_i is bad.

  Next, we want to show that if A is bad and $\epsilon > 0$, then $\exists B \subset A$ such that B is bad and $\operatorname{diam} B < \epsilon$. Indeed, since M is totally bounded, we have a finite $\epsilon/2$-net F. Then $M = \bigcup_{x\in F} B_{\epsilon/2}(x)$, and hence $A = \bigcup_{x\in F}(B_{\epsilon/2}(x) \cap A)$. Since A is bad, there is some $x \in F$ such that $B_{\epsilon/2}(x) \cap A$ is bad, which has a diameter less than ϵ. From this we can construct a sequence $M \supset A_1 \supset A_2 \supset \dots$ such that A_n is bad for any n and $\operatorname{diam} A_n < 1/n$. Picking $x_n \in A_n$ gives a Cauchy sequence x_n which converges to $x \in M$ by completeness. There must be some $U \subset \mathscr{U}$ open that contains x, and it necessarily covers A_n when n is large enough. Then A_n is not bad. Contradiction. $\square$

Remarks.

(i) This is a new proof of Bolzano–Weierstrass theorem. This also proves the Tychonoff's theorem.

(ii) Sadly, the equivalence of sequential compactness and compactness fails in both directions for a general topological space.

7 Differentiation

7.1 Differentiation

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a if

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

exists. So let

$$\epsilon(h) = \frac{f(a+h)-f(a)}{h} - f'(a)\,,$$

then $f(a+h) = f(a) + f'(a)h + \epsilon(h)h$ and $\epsilon \to 0$ as $h \to 0$. We can think of this as $\epsilon(0) = 0$ and ϵ is continuous at 0. We want to use this to define differentiation in higher dimensions. Consider functions from $\mathbb{R}^m$ to $\mathbb{R}^n$, where $m, n \in \mathbb{N}$.

Definition 7.1. $L(\mathbb{R}^m, \mathbb{R}^n)$ denotes the set of linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$.

Definition 7.2. Given $m, n \in \mathbb{N}$ and an open set $U \subset \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and $a \in U$. We say f is *differentiable* at a if there is a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ and a function $\epsilon : \{h \in \mathbb{R}^m \mid a + h \in U\} \to \mathbb{R}^n$ such that

$$f(a+h) = f(a) + T(h) + \epsilon(h)\|h\|\,,$$

where $\epsilon \to 0$ as $h \to 0$ (or equivalently $\epsilon(0) = 0$ and ϵ is continuous at 0).

Remark.

$$\epsilon(h) = \begin{cases} 0 & \text{if } h = 0\,, \\ \dfrac{f(a+h)-f(a)-T(h)}{\|h\|} & \text{if } h \neq 0, a+h \in U\,. \end{cases}$$

Since U is open, $\exists r > 0$ such that $D_r(a) \subset U$, so $D_r(a) \subset \operatorname{Dom}\epsilon$. Note that our condition on ϵ is equivalent to saying $\epsilon(h)\|h\| = o(\|h\|)$ as $h \to 0$.

Alternatively, we may write

$$\lim_{h\to 0} \frac{f(a+h)-f(a)-T(h)}{\|h\|} = 0\,.$$

Next, we observe that such T, if it exists, is unique. If both T, S satisfy our condition, then $(S(h) - T(h))/\|h\| \to 0$ as $h \to 0$, so by choosing $h = x/n$ for $n \in \mathbb{N}$ we have $S = T$.

Definition 7.3. This unique T is called the *derivative* of f at a, usually denoted by $f'(a)$, $Df(a)$ or $Df|_a$, such that

$$f(a+h) = f(a) + f'(a)h + o(\|h\|)\,.$$

Definition 7.4. We say f is *differentiable* on U if it is differentiable at a for every $a \in U$. The *derivative* of f on U is the map

$$\begin{aligned} f' : U &\to L(\mathbb{R}^m, \mathbb{R}^n) \\ x &\mapsto f'(x) \end{aligned}$$

Examples.

(i) Constant functions are differentiable. We can take $f'(a) = 0 \in L(\mathbb{R}^m, \mathbb{R}^n)$.

(ii) Every linear map f is differentiable. Let $f(x) = T(x)$ for some $T \in L(\mathbb{R}^m, \mathbb{R}^n)$. Then

$$f(a+h) = f(a) + f(h) + 0$$

so f is differentiable at a with a derivative T.

(iii) Take $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = \|x\|^2$. Then

$$\begin{aligned} f(a+h) &= \|a+h\|^2 \\ &= \|a\|^2 + 2\langle a|h\rangle + \|h\|^2 \\ &= f(a) + 2\langle a|h\rangle + o(\|h\|)\,, \end{aligned}$$

so $f'(a)(h) = 2\langle a|h\rangle$.

Proposition 7.5 (Differentiable functions are continuous). Let $U \subset \mathbb{R}^m$ and $a \in U$. If $f : U \to \mathbb{R}^n$ is differentiable at a, then f is continuous at a.

Proof. If f is differentiable, then as $h \to 0$, we have

$$f(a+h) - f(a) - f'(a)(h) \to 0\,.$$

Since $f'(a)(h) \to 0$ as well, we must have $f(a+h) \to f(a)$. □

7.2 Operator Norm

So far, we have only looked at derivatives at a single point. We haven't yet discussed much about the derivative at a neighbourhood or the whole space. We might also want to ask if the derivative is continuous or bounded. This is not a straightforward question as we need to define these notions for functions whose values are linear maps. In particular, we want to understand the map $Df : B_r(a) \to L(\mathbb{R}^m, \mathbb{R}^n)$. To do so, we need a metric on the space $L(\mathbb{R}^m, \mathbb{R}^n)$. We will do this by defining a norm.

$L(\mathbb{R}^m, \mathbb{R}^n)$ is a vector space over $\mathbb{R}$ with addition and scalar multiplication defined pointwise. In fact, it is a subspace of $C(\mathbb{R}^m, \mathbb{R}^n)$. To prove this, we will need to show that all linear maps are continuous.

Lemma 7.6. Linear maps in $L(\mathbb{R}^m, \mathbb{R}^n)$ are Lipschitz and hence continuous.

Proof. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^m$ and $x = \sum_i x_i e_i$. By Cauchy–Schwarz,

$$\begin{aligned} \|Tx\| &\le \sum_{i=1}^m |x_i| \|Te_i\| \\ &\le \sqrt{\sum_{i=1}^m |x_i|^2} \sqrt{\sum_{i=1}^m \|Te_i\|^2} \\ &= \|x\| \sqrt{\sum_{i=1}^m \|Te_i\|^2}\,. \end{aligned}$$

□

We can use this fact to define the norm of linear maps. $L(\mathbb{R}^m, \mathbb{R}^n)$ is isomorphic to $\mathbb{R}^{mn}$, both algebraically and topologically, since it is isomorphic to $M_{n\times m}$, the space of $n \times m$ real matrices. It does not really matter which norm we pick as they are all Lipschitz equivalent. We will introduce two of them.

7.2.1 Sup Norm

Definition 7.7. The *sup norm* on $L(\mathbb{R}^m, \mathbb{R}^n)$ is defined by

$$\|A\| = \sup_{x \in \mathbb{R}^m, \|x\|=1} \|Ax\|\,.$$

Thus, $L(\mathbb{R}^m, \mathbb{R}^n)$ becomes a metric space with the metric $d(S,T) = \|S - T\|$.

Proposition 7.8.

(i) $\|A\|$ is finite.

(ii) $\|\cdot\|$ is indeed a norm on $L(\mathbb{R}^m, \mathbb{R}^n)$.

(iii)
$$\|A\| = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|Ax\|}{\|x\|}.$$

(iv) $\|Ax\| \leq \|A\|\|x\|$ for all $x \in \mathbb{R}^m$.

(v) Let $A \in L(\mathbb{R}^m, \mathbb{R}^n)$, $B \in L(\mathbb{R}^n, \mathbb{R}^p)$. Then $BA = B \circ A \in L(\mathbb{R}^m, \mathbb{R}^p)$ and
$$\|BA\| \leq \|B\|\|A\|.$$

Proof.

(i) This is true because A is continuous and $\{x \in \mathbb{R}^m \mid \|x\| = 1\}$ is compact.

(ii) The only non-trivial part is the triangle inequality. We have
$$\begin{aligned}\|A + B\| &= \sup_{\|x\|=1} \|Ax + Bx\| \\ &\leq \sup_{\|x\|=1} (\|Ax\| + \|Bx\|) \\ &\leq \sup_{\|x\|=1} \|Ax\| + \sup_{\|x\|=1} \|Bx\| \\ &= \|A\| + \|B\|.\end{aligned}$$

(iii) This follows from the linearity of A, and the fact that for any $x \in \mathbb{R}^m$, we have
$$\left\| \frac{x}{\|x\|} \right\| = 1.$$

(iv) Immediate from above.

(v)
$$\|BA\| = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|BAx\|}{\|x\|} \leq \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|B\|\|Ax\|}{\|x\|} = \|B\|\|A\|.$$

□

For certain easy cases, we have a straightforward expression for the operator norm.

Proposition 7.9.

(i) If $A \in L(\mathbb{R}, \mathbb{R}^n)$, then A can be written as $Ax = xa$ for some $a \in \mathbb{R}^n$. Moreover, $\|A\| = \|a\|$, where the second norm is the Euclidean norm in $\mathbb{R}^n$.

(ii) If $A \in L(\mathbb{R}^m, \mathbb{R})$, then $Ax = x \cdot a$ for some fixed $a \in \mathbb{R}^m$. Also, $\|A\| = \|a\|$.

Proof.

(i) Set $A(1) = a$. Then by linearity, we get $Ax = xA(1) = xa$. Then we have
$$\|Ax\| = |x|\|a\|.$$

(ii) Let $\{e_i\}$ be the standard basis of $\mathbb{R}^m$. Let $a_i = Ae_i$ and $a = \sum_i a_i e_i$. The rest follows by linearity. □

7.2.2 Euclidean Norm

Let $\{e_i\}_{i=1}^m$ be the standard basis of $\mathbb{R}^m$ and let $\{e'_j\}_{j=1}^n$ be the standard basis of $\mathbb{R}^n$. Then $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ is identified with a $n \times m$ matrix T, where $T_{ji} = \langle Te_i | e'_j \rangle$. We can therefore view $L(\mathbb{R}^m, \mathbb{R}^n)$ as the mn-dimensional vector space $\mathbb{R}^{mn}$, which can be equipped with an Euclidean norm.

Definition 7.10. The *Euclidean norm* on $L(\mathbb{R}^m, \mathbb{R}^n)$ is defined by

$$\|T\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n T_{ji}^2} = \sqrt{\sum_{i=1}^m \|Te_i\|^2},$$

where Te_i is the i-th column of T.

Again, we have the metric $d(S,T) = \|S - T\|$ on $L(\mathbb{R}^m, \mathbb{R}^n)$.

Lemma 7.11.

(i) Given a linear map T, for every $x \in \mathbb{R}^m$, we have $\|Tx\| \le \|T\|\|x\|$.

(ii) For $S \in L(\mathbb{R}^m, \mathbb{R}^n)$, $T \in L(\mathbb{R}^n, \mathbb{R}^p)$, $\|TS\| \le \|T\|\|S\|$.

Proof.

(i) Let $x = \sum_i x_i e_i$, then

$$\|Tx\| \le \sqrt{\sum_{i=1}^m |x_i|^2} \sqrt{\sum_{i=1}^m \|Te_i\|^2} = \|x\|\|T\|$$

by Cauchy–Schwarz, as shown before.

(ii)

$$\|TS\| = \sqrt{\sum_{i=1}^m \|TSe_i\|^2} \le \sqrt{\sum_{i=1}^m \|T\|^2 \|Se_i\|^2} = \|T\|\|S\|.$$

□

As these two definitions of the norm on $L(\mathbb{R}^m, \mathbb{R}^n)$ are equivalent, we will not distinguish them in the future. The notation $\|T\|$ may refer to either of them.

Example. Any bilinear map $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ is differentiable. We have

$$f((a,b) + (h,k)) = f((a+h, b+k)) = f(a,b) + f(a,k) + f(h,b) + f(h,k).$$

Note that $f(a,k) + f(h,b)$ is linear in (h,k), therefore it remains to check $f(h,k) = o(\|(h,k)\|)$. Indeed,

$$\begin{aligned}
\|f(h,k)\| &= \left\| f\left(\sum_{i=1}^m h_i e_i, \sum_{j=1}^n k_j e_j \right) \right\| \\
&\le \sum_{i,j} |h_i||k_j| \|f(e_i, e_j)\| \\
&\le \|(h,k)\|^2 \sum_{i,j} \|f(e_i, e_j)\| \\
&= O(\|(h,k)\|^2) = o(\|(h,k)\|).
\end{aligned}$$

Therefore,

$$f'(a,b)(h,k) = f(a,k) + f(h,b).$$

Let $M_n = L(\mathbb{R}^n, \mathbb{R}^n)$ be the collection of all $n \times n$ real matrices. Consider $f : M_n \to M_n$, $A \mapsto A^2$. It has

$$f(A+H) = A^2 + AH + HA + H^2 = f(A) + AH + HA + o(\|H\|)\,.$$

So $f'(A)(H) = AH + HA$.

Proposition 7.12 (Chain rule). Consider open sets $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ and functions $f : U \to \mathbb{R}^n$, $g : V \to \mathbb{R}^p$ such that $f(U) \subset V$. If f is differentiable at a and g is differentiable at $f(a)$, then $g \circ f$ is differentiable at a and

$$(g \circ f)'(a) = g'(f(a))f'(a)\,.$$

Proof. Let $A = f'(a)$ and $B = g'(f(a))$. By the differentiability of f, we know

$$\begin{cases} f(a+h) = f(a) + Ah + o(\|h\|) \\ g(f(a)+k) = g(f(a)) + Bk + o(\|k\|)\,. \end{cases}$$

Now we have

$$\begin{aligned} g \circ f(a+h) &= g(f(a) + \underbrace{Ah + o(\|h\|)}_{k}) \\ &= g(f(a)) + B(Ah + o(\|h\|)) + o(Ah + o(\|h\|)) \\ &= g \circ f(a) + BAh + B(o(\|h\|)) + o(Ah + o(\|h\|))\,. \end{aligned}$$

We only have to show that the last two terms are $o(\|h\|)$. This is true since

$$B(o(\|h\|)) \le \|B\|\|o(\|h\|)\| = o(\|h\|)\,,$$

and for sufficiently small h,

$$\|Ah + o(\|h\|)\| \le \|A\|\|h\| + \|o(\|h\|)\| \le (\|A\| + 1)\|h\|\,,$$

so $o(Ah + o(\|h\|))$ is $o(\|h\|)$ as well. Hence

$$g \circ f(a+h) = g \circ f(a) + BAh + o(\|h\|)\,.$$

□

Proposition 7.13. Let $U \subset \mathbb{R}^m$ be open. $f : U \to \mathbb{R}^n$ is differentiable at $a \in U$ if and only if each component $f_j = \pi_j \circ f$ is differentiable at a. Also,

$$f'(a)(h) = \sum_{j=1}^{n} f_j'(a)(h)e_j'\,.$$

Proof.

($\Rightarrow$) $\pi_j(x) = \langle x | e_j' \rangle$ is linear hence differentiable. Trivial by chain rule.

($\Leftarrow$)

$$\begin{aligned} f(a+h) &= \sum_{j=1}^{n} f_j(a+h)e_j' \\ &= \sum_{j=1}^{n} (f_j(a) + f_j'(a)(h) + \epsilon_j(h)\|h\|)e_j' \\ &= f(a) + \left(\sum_{j=1}^{n} f_j'(a)(h)e_j'\right) + \left(\sum_{j=1}^{n} \epsilon_j(h)e_j'\right)\|h\|\,, \end{aligned}$$

where the last term is $o(\|h\|)$. □

Proposition 7.14 (Linearity and product rule). Let $f, g : U \to \mathbb{R}^n$, $\phi : U \to \mathbb{R}$, where $U \subset \mathbb{R}^m$ is open. If f, g, ϕ are differentiable at $a \in U$, then so are $f + g$ and ϕf, and

$$(f+g)'(a) = f'(a) + g'(a)\,,$$
$$(\phi f)'(a)(h) = \phi'(a)(h) f(a) + \phi(a) f'(a)(h)\,.$$

Proof. Have

$$f(a+h) = f(a) + f'(a)(h) + o(\|h\|)\,,$$
$$g(a+h) = g(a) + g'(a)(h) + o(\|h\|)\,,$$
$$\phi(a+h) = \phi(a) + \phi'(a)(h) + o(\|h\|)\,.$$

Hence

$$(f+g)(a+h) = (f+g)(a) + (f'(a) + g'(a))(h) + o(\|h\|)\,.$$

We can do the same thing to derive the product rule, but here we will provide an alternative proof. Let $F : U \to \mathbb{R} \times \mathbb{R}^n = \mathbb{R}^{n+1}$ by $x \mapsto (\phi(x), f(x))$, and $G : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ by $(a, x) \mapsto ax$. F is differentiable by the previous proposition, and G is differentiable since it is bilinear. Hence $\phi f = G \circ F$ is differentiable and we can obtain the form of the derivative by chain rule. □

7.3 Directional and Partial Derivatives

While the definition of the derivative is good, it is purely existential. This is unlike the definition of differentiability of real functions, where we are asked to compute an explicit limit if the limit exists, that's the derivative. If not, it is not differentiable. In the higher-dimensional world, this is not the case. We have completely no idea where to find the derivative, even if we know it exists. So we would like an explicit formula for it.

The idea is to look at specific directions instead of finding the general derivative. As always, let $f : U \to \mathbb{R}^n$ be differentiable at $a \in U$. Fix some non-zero $u \in \mathbb{R}^m$, take $h = tu$ with $t \in \mathbb{R}$. Differentiability tells

$$\lim_{t\to 0} \frac{f(a+tu) - f(a) - f'(a)(tu)}{\|tu\|} = 0$$
$$\iff \lim_{t\to 0} \frac{f(a+tu) - f(a) - tf'(a)(u)}{t} = 0$$
$$\iff f'(a)(u) = \lim_{t\to 0} \frac{f(a+tu) - f(a)}{t}\,.$$

We derived this assuming $u \neq 0$, but this is trivially true for $u = 0$. So this is valid for all $u \in \mathbb{R}^m$.

This is of the same form as the usual derivative, and it is usually not too difficult to compute this limit. Note, however, that this says if the derivative exists, then the limit above is related to the derivative as above. However, even if the limit exists for all u, we still cannot conclude that the derivative exists. Regardless, even if the derivative does not exist, this limit is still often a useful notion.

Definition 7.15. Let $f : U \to \mathbb{R}^n$, where $U \subset \mathbb{R}^m$ is open. For $u \in \mathbb{R}^m$ and $a \in U$, the *directional derivative* of f at a in the direction of u is the limit

$$D_u f(a) \coloneqq \lim_{t\to 0} \frac{f(a+tu) - f(a)}{t}\,.$$

By definition, we have

$$D_u f(a) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(a+tu)\,.$$

Often, it is convenient to focus on the special cases where $u = e_i$, a member of the standard basis for $\mathbb{R}^m$. This is known as the partial derivative. By convention, this is defined for real-valued functions only, but the same definition works for any $\mathbb{R}^n$-valued function.

Definition 7.16. The j-th partial derivative of $f : U \to \mathbb{R}$ at $a \in U \subset \mathbb{R}^m$ is

$$D_{e_j} f(a) = \lim_{t \to 0} \frac{f(a + te_j) - f(a)}{t}$$

if the limit exists. This is often written

$$D_{e_j} f(a) = D_j f(a) = \frac{\partial f}{\partial x_j}\,.$$

Note that these definitions do not require differentiability of f at a. We will see some examples shortly. Before that, we first establish some elementary properties of differentiable functions.

Proposition 7.17. Let $U \subset \mathbb{R}^m$ be open and $a \in U$.

(i) If $f : U \to \mathbb{R}^n$ is differentiable at a, then the directional derivative $D_u f(a)$ exists for any $u \in \mathbb{R}^m$, and
$$D_u f(a) = Df(a)u\,.$$

(ii) If $f = (f_1, \dots, f_n) : U \to \mathbb{R}^n$ is differentiable at a, then all partial derivatives $D_j f_i(a)$ exist for all $i = 1, \dots, n$, $j = 1, \dots, m$, and are given by
$$D_j f_i(a) = Df_i(a)e_j\,.$$

(iii) If A is the matrix representing $Df(a)$ with respect to the basis for $\mathbb{R}^m$ and $\mathbb{R}^n$, i.e. for any $h \in \mathbb{R}^m$,
$$Df(a)h = \mathsf{A}h\,.$$
Then A is given by
$$A_{ij} = \langle Df(a)e_j | b_i \rangle = D_j f_i(a)\,,$$
where $\{e_1, \dots, e_m\}$ and $\{b_1, \dots, b_n\}$ are the standard basis of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.

Remark. This is often expressed in the form

$$f'(a)(h) = \sum_{i=1}^{m} h_i D_i f(a)\,.$$

Proof.

(i) Obvious from above discussion.

(ii) Obvious from above discussion.

(iii) This follows from the general result for linear maps: for any linear map represented by a $m \times n$ matrix (A_{ij}), we have
$$A_{ij} = \langle \mathsf{A}e_j | b \rangle\ .$$
Apply this with $\mathsf{A} = Df(a)$ and note that for any $h \in \mathbb{R}^n$,
$$Df(a)(h) = (Df_1(a)h, \dots, Df_m(a)h)\,.$$
So done. □

Definition 7.18. Let f be differentiable at a. The *Jacobian matrix* of f at a, denoted $J_f(a)$ is the matrix of $f'(a)$ with respect to the standard bases

$$(J_f(a))_{ji} = \left\langle D_i f(a) \middle| e_j' \right\rangle = \pi_j(D_i f(a)) = D_i f_j(a) = \frac{\partial f_j}{\partial x_i}\,.$$

The above says differentiability at a point implies the existence of all directional derivatives, which in turn implies the existence of all partial derivatives. The converse implication does not hold in either of these.

Example. Define

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto \begin{cases} 0 & xy = 0 \\ 1 & xy \neq 0\,. \end{cases} \end{aligned}$$

Then the partial derivatives at $(0, 0)$ are

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0\,.$$

In other directions, say $u = (1, 1)$, we have

$$D_u f(0, 0) = \lim_{t\to 0} \frac{f((0, 0) + tu) - f(0, 0)}{t} = \lim_{t\to 0} \frac{1}{t}$$

which diverges. So the directional derivative does not exist.

Example. Define

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto \begin{cases} \dfrac{x^3}{y} & y \neq 0 \\ 0 & y = 0\,. \end{cases} \end{aligned}$$

Then for $u = (u_1, u_2) \neq 0$ and $t \neq 0$, we can compute

$$\frac{f((0, 0) + tu) - f(0, 0)}{t} = \begin{cases} \dfrac{tu_1^3}{u_2} & u_2 \neq 0 \\ 0 & u_2 = 0\,. \end{cases}$$

So

$$D_u f(0, 0) = \lim_{t\to 0} \frac{f((0, 0) + tu) - f(0, 0)}{t} = 0\,,$$

and the directional derivatives exist in all directions. However, the function is not differentiable at $(0, 0)$, since it is even not continuous there.

Example. Define

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto \begin{cases} \dfrac{x^3}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0)\,. \end{cases} \end{aligned}$$

It is clear that f is continuous at points other than $(0, 0)$. f is also continuous at $(0, 0)$ since $|f(x, y)| \leq |x|$. We can compute the partial derivatives as

$$\frac{\partial f}{\partial x} = 1\,,\ \frac{\partial f}{\partial y} = 0\,.$$

In fact, we can compute the difference quotient in the direction $u = (u_1, u_2) \neq (0, 0)$ to be

$$\frac{f((0, 0) + tu) - f(0, 0)}{t} = \frac{u_1^3}{u_1^2 + u_2^2}\,.$$

So we have

$$D_u f(0,0) = \frac{u_1^3}{u_1^2 + u_2^2}\,.$$

We can now immediately conclude that f is not differentiable at $(0,0)$, since if it were, then we would have

$$D_u f(0,0) = Df(0,0)u\,,$$

which should be linear in u, but this is not.

Alternatively, if f were differentiable, then we have

$$Df(0,0)h = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1\,.$$

However, we have

$$\frac{f((0,0)+h) - f(0,0) - Df(0,0)h}{\|h\|} = -\frac{h_1 h_2^2}{\sqrt{h_1^2 + h_2^2}^3}\,,$$

which does not tend to 0 as $h \to (0,0)$. For example, if $h = (t,t)$, this quotient is $-2^{-3/2}$ for $t \neq 0$.

To decide if a function is differentiable, the first step would be to compute the partial derivatives. If they do not exist, then we can immediately know the function is not differentiable. However, if they do, we then have a candidate for the derivative, and we can plug it into the definition to check that if it is actually the derivative.

This is a cumbersome thing to do. It turns out that while the existence of partial derivatives does not imply differentiability in general, we can get differentiability if we add some slight conditions.

Theorem 7.19. Let $U \subset \mathbb{R}^m$ be open, $f : U \to \mathbb{R}^n$. Let $a \in U$. Suppose there is some open ball $D_r(a) \subset U$ such that

(i) $D_j f_i(x)$ exists for every $x \in D_r(a)$, $1 \leq i \leq n$, $1 \leq j \leq m$;

(ii) $D_j f_i$ are continuous at a for all $1 \leq i \leq n$, $1 \leq j \leq m$.

Then f is differentiable at a.

Proof. We only need to prove for the $n = 1$ case, as for $n > 1$, we can break f into components. For each $h = (h_1, \dots, h_m) \in \mathbb{R}^m$, we have

$$f(a+h) - f(a) = \sum_{j=1}^{m} f(a + h_1 e_1 + \cdots + h_j e_j) - f(a + h_1 e_1 + \cdots + h_{j-1} e_{j-1})\,.$$

For convenience, write

$$h^{(j)} = h_1 e_1 + \cdots + h_j e_j = (h_1, \dots, h_j, 0, \dots, 0)\,.$$

Then we have

$$\begin{aligned} f(a+h) - f(a) &= \sum_{j=1}^{m} f(a + h^{(j)}) - f(a + h^{(j-1)}) \\ &= \sum_{j=1}^{m} f(a + h^{(j-1)} + h_j e_j) - f(a + h^{(j-1)})\,. \end{aligned}$$

Note that in each term, we are just moving along the coordinate axes. Since the partial derivatives exist, the mean value theorem of single-variable calculus applies to

$$g(t) = f(a + h^{(j-1)} + t h_j e_j)$$

on the interval $t \in [0, 1]$ allows us to write this as

$$\begin{aligned} f(a+h) - f(a) &= \sum_{j=1}^{m} h_j D_j f(a + h^{(j-1)} + \theta_j h_j e_j) \\ &= \sum_{j=1}^{m} h_j D_j f(a) + \sum_{j=1}^{m} h_j \Big(D_j f(a + h^{(j-1)} + \theta_j h_j e_j) - D_j f(a) \Big) \end{aligned}$$

for some $\theta_j \in (0, 1)$.

Note that $D_j f(a + h^{(j-1)} + \theta_j h_j e_j) - D_j f(a) \to 0$ as $h \to 0$ since the partial derivatives are continuous at a. So the second term is $o(\|h\|)$, so f is differentiable at a with

$$Df(a)h = \sum_{j=1}^{m} D_j f(a) h_j \, .$$

□

Example. As a random example,

$$\begin{aligned} f : \mathbb{R}^3 &\to \mathbb{R}^2 \\ (x, y, z) &\mapsto (3x^2 + 4 \sin y + \mathrm{e}^{6z}, xyz\mathrm{e}^{14x}) \end{aligned}$$

is differentiable everywhere since it has continuous partial derivatives.

7.4 Mean Value Inequalities

So far, we have just looked at cases where we assume the function is differentiable at a point. We are now going to assume the function is differentiable in a region, and see what happens to the derivative.

Recall the mean value theorem for single-variable calculus: if $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then

$$f(b) - f(a) = f'(c)(b - a)$$

for some $c \in (a, b)$. Here we have an exact equality. However, in general, for vector-valued functions, this is no longer true. Instead, we only have an inequality.

We will first prove the case when the domain is a subset of $\mathbb{R}$.

Theorem 7.20. Let $f : [a, b] \to \mathbb{R}^n$ be continuous on $[a, b]$ and differentiable on (a, b). Suppose there is some M such that for all $t \in (a, b)$, we have $\|Df(t)\| \leq M$, then

$$\|f(b) - f(a)\| \leq M(b - a) \, .$$

Proof. Let $v = f(b) - f(a)$. Define

$$g(t) = v \cdot f(t) = \sum_{i=1}^{n} v_i f_i(t) \, .$$

Since each f_i is differentiable, g is continuous on $[a, b]$ and differentiable on (a, b) with

$$g'(t) = \sum v_i f_i'(t) \, .$$

Hence, we know

$$|g'(t)| \leq \left| \sum_{i=1}^{n} v_i f_i'(t) \right| \leq \|v\| \left(\sum_{i=1}^{n} f_i'^2(t) \right)^{1/2} = \|v\| \|Df(t)\| \leq M\|v\| \, .$$

We now apply the mean value theorem to g to get

$$g(b) - g(a) = g'(t)(b-a)$$

for some $t \in (a, b)$. By definition of g, we get

$$v \cdot (f(b) - f(a)) = g'(t)(b-a)\,.$$

By the definition of v, we have

$$\|f(b) - f(a)\|^2 = |g'(t)(b-a)| \leq (b-a)M\|f(b) - f(a)\|\,.$$

If $f(b) = f(a)$, then there is nothing to prove. Otherwise, divide by $\|f(b) - f(a)\|$ and done. □

We can now apply this to prove the general version.

Theorem 7.21 (Mean value inequality). Let $a \in \mathbb{R}^m$ and $f : D_r(a) \to \mathbb{R}^n$ be differentiable on $D_r(a)$ with $\|Df(x)\| \leq M$ for all $x \in D_r(a)$. Then

$$\|f(b_1) - f(b_2)\| \leq M\|b_1 - b_2\|$$

for any $b_1, b_2 \in D_r(a)$.

Proof. We will reduce it to the previous $m = 1$ case.

Fix $b_1, b_2 \in D_r(a)$. Note that

$$tb_1 + (1-t)b_2 \in D_r(a)$$

for all $t \in [0, 1]$. Now consider

$$\begin{aligned} g : [0,1] &\to \mathbb{R}^n \\ t &\mapsto f(tb_1 + (1-t)b_2)\,. \end{aligned}$$

By the chain rule, g is differentiable and

$$g'(t) = Dg(t) = (Df(tb_1 + (1-t)b_2))(b_1 - b_2)\,.$$

Therefore,

$$\|Dg(t)\| \leq \|Df(tb_1 + (1-t)b_2)\|\|b_1 - b_2\| \leq M\|b_1 - b_2\|\,.$$

Now we can apply the previous theorem and get

$$\|f(b_1) - f(b_2)\| = \|g(1) - g(0)\| \leq M\|b_1 - b_2\|\,.$$

□

Note that here we worked in a ball. In general, we could have worked in any convex set, since all we need is for $tb_1 + (1-t)b_2$ to be inside the domain.

But with this, we have the following easy corollary.

Corollary. Let $f : D_r(a) \subset \mathbb{R}^m \to \mathbb{R}^n$ have $Df(x) = 0$ for all $x \in D_r(a)$. Then f is constant.

Proof. Trivial. □

We would like to extend this corollary. Does this corollary extend to differentiable maps f with $Df = 0$ defined on any open set $U \subset \mathbb{R}^m$?

The answer is clearly no, even for $f : \mathbb{R} \to \mathbb{R}$. Consider f defined on disjoint intervals $[1, 2] \cup [3, 4]$, where we define $f(t)$ to be 1 on $[1, 2]$ and 2 on $[3, 4]$. Then $Df = 0$ but f is clearly not a constant. It is just locally constant on each interval.

The problem is that the sets are disconnected. We cannot connect points in $[1, 2]$ and points in $[3, 4]$ with a line. Recall the notion of path-connectedness.

Theorem 7.22. Let $U \subset \mathbb{R}^m$ be open and path-connected. Then for any differentiable $f : U \to \mathbb{R}^n$, if $Df(x) = 0$ for all $x \in U$, then f is constant on U.

A naive attempt would be to replace $tb_1 + (1-t)b_2$ in the proof of the mean value theorem with a path $\gamma : [0,1] \to \mathbb{R}^n$. However, this is not a correct proof, since this has to assume γ is differentiable. So this does not work in general. We have to think some more.

Proof. We are going to use the fact that f is locally constant. Wlog, let $n = 1$. Given any $a, b \in U$, we need to show that $f(a) = f(b)$. Let $\gamma : [0,1] \to U$ be a continuous path from a to b. For any $s \in (0,1)$, there exists some ϵ such that $D_\epsilon(\gamma(s)) \subset U$ since U is open. By continuity of γ, there is a δ such that $(s-\delta, s+\delta) \subset [0,1]$ with $\gamma((s-\delta, s+\delta)) \subset D_\epsilon(\gamma(s)) \subset U$.

Since f is constant on $D_\epsilon(\gamma(s))$ by the previous corollary, we know that $g(t) = f \circ \gamma(t)$ is constant on $(s-\delta, s+\delta)$. In particular, g is differentiable at s with derivative 0. This is true for all s. So the map $g : [0,1] \to \mathbb{R}$ has zero derivative on $(0,1)$ and is continuous on $(0,1)$. So g is a constant. Then $g(0) = g(1)$, i.e. $f(a) = f(b)$. □

If γ were differentiable, then this is much easier, since we can show $g' = 0$ by the chain rule:

$$g'(t) = Df(\gamma(t))\gamma'(t)\,.$$

7.5 Inverse Function Theorem

Definition 7.23. Let $U \subset \mathbb{R}^m$ be open. We say $f : U \to \mathbb{R}^n$ is C^1 on U if f is differentiable at each $x \in U$ and

$$Df : U \to L(\mathbb{R}^m, \mathbb{R}^n)$$

is continuous.

We write $C^1(U)$ or $C^1(U; \mathbb{R}^n)$ for the set of all C^1 maps from U to $\mathbb{R}^n$.

First, we get a convenient alternative characterisation of C^1.

Proposition 7.24. Let $U \subset \mathbb{R}^m$ be open. Then $f = (f_1, \dots, f_n) : U \to \mathbb{R}^n$ is C^1 on U if and only if the partial derivatives $D_j f_i(x)$ exist for all $x \in U$, $1 \le i \le n$, $1 \le j \le m$ and $D_j f_i : U \to \mathbb{R}$ are continuous.

Proof. ($\Rightarrow$) Differentiability of f at x implies $D_j f_i(x)$ exists and are given by

$$D_j f_i(x) = \langle Df(x)e_j | b_i \rangle\ ,$$

where $\{e_1, \dots, e_m\}$ and $\{b_1, \dots, b_n\}$ are the standard bases for $\mathbb{R}^m$ and $\mathbb{R}^n$. So we know

$$|D_j f_i(x) - D_j f_i(y)| = |\langle (Df(x) - Df(y))e_j | b_i \rangle| \le \|Df(x) - Df(y)\|$$

since e_j and b_i are unit vectors. Hence if Df is continuous, so is $D_j f_i$.

($\Leftarrow$) Since the partials exist and are continuous, by the previous theorem, we know that the derivative Df exists. To show $Df : U \to L(\mathbb{R}^m, \mathbb{R}^n)$ is continuous, recall

$$\|A\| = \sqrt{\sum\sum A_{ij}^2}\,.$$

Apply this to $A = Df(x) - Df(y)$, get

$$\|Df(x) - Df(y)\| = \sqrt{\sum\sum (D_j f_i(x) - D_j f_i(y))^2}\,.$$

So if all $D_j f_i$ are continuous, then so is Df. □

Finally, we can go to the inverse function theorem.

Theorem 7.25 (Inverse function theorem). Let $U \subset \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^n$ be a C^1 map. Let $a \in U$ and suppose that $Df(a)$ is invertible as a linear map $\mathbb{R}^n \to \mathbb{R}^n$. Then there exists open sets $V, W \subset \mathbb{R}^n$ with $a \in V$, $f(a) \in W$, $V \subset U$ such that

$$f|_V : V \to W$$

is a bijection. Moreover, the inverse map $f|_V^{-1} : W \to V$ is also C^1.

We have a name for these functions.

Definition 7.26. Let $U, U' \subset \mathbb{R}^n$ be open. A map $g : U \to U'$ is a *diffeomorphism* if it is C^1 with a C^1 inverse.

Then the inverse function theorem states that if f is C^1 and $Df(a)$ is invertible, then f is a local diffeomorphism at a.

Before proving this, let us first look at the simple case where $n = 1$. Suppose that $f'(a) \neq 0$, then there exists a δ such that $f'(t) > 0$ or $f'(t) < 0$ in $t \in (a - \delta, a + \delta)$. So $f|_{(a-\delta,a+\delta)}$ is monotone and hence is invertible. This is trivial. However, this is not trivial for $n \geq 2$.

Proof. By replacing f with $(Df(a))^{-1}f$ (or by rotating our heads and stretching it a little bit), we can assume $Df(a) = I$, the identity map. By the continuity of Df, there exists some $r > 0$ such that for all $x \in B_r(a)$,

$$\|Df(x) - I\| < \frac{1}{2}.$$

By shrinking r sufficiently, we can assume $B_r(a) \subset U$. Let $W = D_{r/2}(f(a))$, and let $V = f^{-1}(W) \cap D_r(a)$.

This is our setup. We need three steps to prove this theorem.

Claim 1. V is open, and $f|_V : V \to W$ is a bijection.

Since f is continuous, $f^{-1}(W)$ is open. So V is open. To show $f|_V : V \to W$ is a bijection, we have to show that for each $y \in W$, there is a unique $x \in V$ such that $f(x) = y$. We are going to use the contraction mapping theorem to prove this. This statement is equivalent to proving that for each $y \in W$, the map $T(x) = x - f(x) + y$ has a unique fixed point $x \in V$.

Let $h(x) = x - f(x)$. Then note that

$$Dh(x) = I - Df(x).$$

So by our choice of r, for every $x \in B_r(a)$, we must have

$$\|Dh(x)\| \leq \frac{1}{2}.$$

Then for any $x_1, x_2 \in B_r(a)$, we can use the mean value inequality to estimate

$$\|h(x_1) - h(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

Hence we know

$$\|T(x_1) - T(x_2)\| = \|h(x_1) - h(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

Finally, to apply the contraction mapping theorem, we need to pick the right domain for T, namely $B_r(a)$. For any $x \in B_r(a)$, we have

$$\begin{aligned}\|T(x) - a\| &= \|x - f(x) + y - a\| \\ &= \|x - f(x) - (a - f(a)) + y - f(a)\| \\ &\leq \|h(x) - h(a)\| + \|y - f(a)\| \\ &\leq \frac{1}{2}\|x - a\| + \|y - f(a)\| \\ &< \frac{r}{2} + \frac{r}{2} = r\,.\end{aligned}$$

So $T : B_r(a) \to D_r(a) \subset B_r(a)$. Since $B_r(a)$ is complete, T has a unique fixed point $x \in B_r(a)$, i.e. $T(x) = x$. Finally, we need to show $x \in D_r(a)$, since this is where we want to find our fixed point. This is true since $T(x) \in D_r(a)$ by above, so we must have $x \in D_r(a)$. Also, since $f(x) = y$, we know $x \in f^{-1}(W)$ so $x \in V$. This finishes the proof of our first claim.

It remains to show that $f|_V$ is invertible with a C^1 inverse.

Claim 2. The inverse map $g = f|_V^{-1} : W \to V$ is Lipschitz and hence continuous. In fact, we have

$$\|g(y_1) - g(y_2)\| \leq 2\|y_1 - y_2\|\,.$$

For any $x_1, x_2 \in V$, by the triangle inequality, we have

$$\begin{aligned}\|x_1 - x_2\| - \|f(x_1) - f(x_2)\| &\leq \|(x_1 - f(x_1)) - (x_2 - f(x_2))\| \\ &= \|h(x_1) - h(x_2)\| \\ &\leq \frac{1}{2}\|x_1 - x_2\|\,.\end{aligned}$$

Hence, we get

$$\|x_1 - x_2\| \leq 2\|f(x_1) - f(x_2)\|\,.$$

Apply this to $x_1 = g(y_1)$ and $x_2 = g(y_2)$, and note that $f(g(y_i)) = y_i$ to get the claimed result.

Claim 3. g is C^1, and for all $y \in W$,

$$Dg(y) = Df(g(y))^{-1}\,. \tag{$\dagger$}$$

Note that if g were differentiable, then its derivative must be given by ($\dagger$), since by definition, we know

$$f(g(y)) = y\,,$$

and hence the chain rule gives

$$Df(g(y)) \cdot Dg(y) = I\,.$$

Also, we immediately know Dg is continuous, since it is the composition of continuous functions (the inverse of a matrix is given by a polynomial expression of its components). So it suffices to check that $Df(g(y))^{-1}$ satisfies the definition of the derivative.

First we check that $Df(x)$ is indeed invertible for every $x \in B_r(a)$. We use the fact that

$$\|Df(x) - I\| \leq \frac{1}{2}\,.$$

If $Df(x)v = 0$, then we have

$$\|v\| = \|Df(x)v - v\| \leq \|Df(x) - I\|\|v\| \leq \frac{1}{2}\|v\|\,,$$

so we must have $\|v\| = 0 \implies v = 0$. So $\ker Df(x) = \{0\}$. So $Df(g(y))^{-1}$ exists.

Let $x \in V$ be fixed, and $y = f(x)$. Let k be small and

$$h = g(y + k) - g(y)\,,$$

or in other words,

$$f(x + h) - f(x) = k\,.$$

Since g is invertible, whenever $k \neq 0$, we have $h \neq 0$. Since g is continuous, as $k \to 0$, $h \to 0$ as well. We have

$$\begin{aligned}
&\frac{g(y+k) - g(y) - Df(g(y))^{-1}k}{\|k\|}\\
&= \frac{h - Df(g(y))^{-1}k}{\|k\|}\\
&= \frac{Df(x)^{-1}(Df(x)h - k)}{\|k\|}\\
&= \frac{-Df(x)^{-1}(f(x+h) - f(x) - Df(x)h)}{\|k\|}\\
&= -Df(x)^{-1}\left(\frac{f(x+h) - f(x) - Df(x)h}{\|h\|}\frac{\|h\|}{\|k\|}\right)\\
&= -Df(x)^{-1}\left(\frac{f(x+h) - f(x) - Df(x)h}{\|h\|}\frac{\|g(y+k) - g(y)\|}{\|(y+k) - y\|}\right).
\end{aligned}$$

As $k \to 0$, $h \to 0$. The first factor $-Df(x)^{-1}$ is fixed; the second factor tends to 0 as $h \to 0$; the third factor is bounded by 2. The whole thing tends to 0, so done. □

Remark. Note that in the case where $n = 1$, if $f : (a, b) \to \mathbb{R}$ is C^1 with $f'(x) \neq 0$ for every x, then f is monotone on the whole domain (a, b), and hence $f : (a, b) \to f((a, b))$ is a bijection. In higher dimensions, this is not true. Even if we know that $Df(x)$ is invertible for all $x \in U$, we cannot guarantee that $f|_U$ is a bijection. We only know that there is a local inverse.

Example. Let $U = \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(x, y) = (\mathrm{e}^x \cos y, \mathrm{e}^x \sin y)\,.$$

Then we can directly compute

$$Df(x, y) = \begin{pmatrix} \mathrm{e}^x \cos y & -\mathrm{e}^x \sin y \\ \mathrm{e}^x \sin y & \mathrm{e}^x \cos y \end{pmatrix}.$$

Then we have

$$\det(Df(x, y)) = \mathrm{e}^{2x} \neq 0$$

for all $(x, y) \in \mathbb{R}^2$. However, by periodicity, we have

$$f(x, y + 2n\pi) = f(x, y)$$

for all $n \in \mathbb{Z}$. So f is not injective on $\mathbb{R}^2$.

One major application of the inverse function theorem is to prove the implicit function theorem, the statement of which is as follows.

Theorem 7.27 (Implicit function theorem). Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m \in C^1$, and let $\mathbb{R}^{n+m}$ have coordinates (x, y). Fix a point $(a, b) = (a_1, \ldots, a_n, b_1, \ldots, b_m)$ with $f(a, b) = 0$. If the Jacobian matrix with respect to y

$$J_{f,y}(a, b) = \left(\frac{\partial f_i}{\partial y_j}\right)$$

is invertible, then there exists an open set $U \subset \mathbb{R}^n$ containing a such that there exists a unique function $g : U \to \mathbb{R}^m$ such that $g(a) = b$ and $f(x, g(x)) = 0$ for all $x \in U$. Moreover, $g \in C^1$ and denoting the Jacobian with respect to x as

$$J_{f,x}(a, b) = \left(\frac{\partial f_i}{\partial x_j}\right),$$

the Jacobian matrix of partial derivatives of g in U is given by the matrix product

$$\left(\frac{\partial g_i}{\partial x_j}\right)_{m\times n} = -(J_{f,y}(x, g(x)))^{-1}_{m\times m}(J_{f,x}(x, g(x)))_{m\times n}.$$

The proof is really intricate so we will not do anything further.

7.6 Second Order Derivative

For real functions, we can immediately obtain the notion of higher derivatives, since the derivative is just a normal function again. Here, things are slightly more complicated, since the derivative is a linear operator. However, this is not too much of a problem since the space of linear operators are just yet another vector space. We can use essentially the same definition.

Definition 7.28. Let $U \subset \mathbb{R}^m$ be open, $f : U \to \mathbb{R}^n$ be differentiable, with derivative $Df : U \to L(\mathbb{R}^m, \mathbb{R}^n)$. We say Df is differentiable at $a \in U$ if there exists $A \in L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$ such that

$$\lim_{h\to 0} \frac{1}{\|h\|}(Df(a + h) - Df(a) - Ah) = 0.$$

For this to make sense, we need a norm on $L(\mathbb{R}^m, \mathbb{R}^n)$ (e.g. the sup norm), but such an A, if exists, is independent of the choice of the norm as all norms are equivalent for a finite-dimensional space.

This is in fact the same definition as our usual differentiability, since $L(\mathbb{R}^m, \mathbb{R}^n)$ is a finite-dimensional vector space isomorphic to $\mathbb{R}^{mn}$. So Df is differentiable if and only if $Df : U \to \mathbb{R}^{mn}$ is differentiable with $A \in L(\mathbb{R}^m, \mathbb{R}^{mn})$. This allows us to use our previous definitions and theorems about differentiability.

Proposition 7.29. Let $U \subset \mathbb{R}^m$ open and $f : U \to \mathbb{R}^n$ be differentiable. Df is differentiable at $a \in U$ if and only if the partial derivatives $D_i(D_j f_k)$ exist in a neighbourhood of a and are continuous at a for all $i, j = 1, \dots, m$ and $k = 1, \dots, n$.

Proof. Trivial. □

Definition 7.30. We denote the *second-order partial derivatives*

$$D_{ij}f(a) = D_i(D_j f)(a) = \frac{\partial^2}{\partial x_i \partial x_j} f(a).$$

Let us now go back to the initial definition and try to interpret it. By linear algebra, in general, a linear map $\phi : \mathbb{R}^l \to L(\mathbb{R}^m, \mathbb{R}^n)$ induces a bilinear map $\Phi : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Phi(u, v) = \phi(u)(v) \in \mathbb{R}^n.$$

In particular, we know

$$\Phi(au + bv, w) = a\Phi(u, w) + b\Phi(v, w)$$
$$\Phi(u, av + bw) = a\Phi(u, v) + b\Phi(u, w).$$

Conversely, if $\Phi : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ is bilinear, then $\phi : \mathbb{R}^l \to L(\mathbb{R}^m, \mathbb{R}^n)$ defined by

$$\phi(u) = (v \mapsto \Phi(u, v))$$

is linear. These are clearly inverse operations to each other. So there is a one-to-one correspondence between bilinear maps $\Phi : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ and linear maps $\phi : \mathbb{R}^l \to L(\mathbb{R}^m, \mathbb{R}^n)$.

Instead of treating our second derivative as a weird linear map in $L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$, we can view it as a bilinear map $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$.

Definition 7.31. For $f : U \to \mathbb{R}^n$ twice differentiable, we define the *second derivative* at $a \in U$ as the bilinear map

$$\begin{aligned} D^2 f(a) : \mathbb{R}^m \times \mathbb{R}^m &\to \mathbb{R}^n \\ (u, v) &\mapsto D(Df)(a)(u)(v)\,. \end{aligned}$$

In coordinates, if we define

$$u = \sum_{i=1}^m u_i e_i\,,\ v = \sum_{j=1}^m v_i e_i\,,$$

where $\{e_1, \dots, e_m\}$ is the standard basis for $\mathbb{R}^m$, then using the bilinearity, we have

$$D^2 f(a)(u, v) = \sum_{i=1}^m \sum_{j=1}^m D^2 f(a)(e_i, e_j) u_i v_j\,.$$

This is very similar to the case of first derivatives, where the derivative can be completely specified by the values it takes on the basis vectors.

In the definition of the second derivative, we can again take $h = te_i$. Then

$$\lim_{t \to 0} \frac{Df(a + te_i) - Df(a) - tD(Df)(a)(e_i)}{t} = 0\,.$$

Note that the whole thing at the top is a linear map in $L(\mathbb{R}^m, \mathbb{R}^n)$. We can then act the whole thing on e_j and obtain

$$\lim_{t \to 0} \frac{Df(a + te_i)(e_j) - Df(a)(e_j) - tD(Df)(a)(e_i)(e_j)}{t} = 0$$

for all $i, j = 1, \dots, n$. By rearranging, we have

$$\begin{aligned} D^2 f(a)(e_i, e_j) &= \lim_{t \to 0} \frac{Df(a + te_i)(e_j) - Df(a)(e_j)}{t} \\ &= \lim_{t \to 0} \frac{D_j f(a + te_i) - D_j f(a)}{t} \\ &= D_i D_j f(a)\,. \end{aligned}$$

Hence, we have

$$D^2 f(a)(e_i, e_j) = \sum_{k=1}^n D_{ij} f_k(a) b_k\,,$$

where $\{b_1, \dots, b_n\}$ is the standard basis for $\mathbb{R}^n$. So

$$D^2 f(a)(u, v) = \sum_{i,j=1}^m \sum_{k=1}^n D_{ij} f_k(a) u_i v_j b_k\,.$$

So far, we have been very careful about keeping the right order of the partial derivatives. However, in most cases we care about, this does not really matter.

Theorem 7.32 (Symmetry of mixed partials). Let $U \subset \mathbb{R}^m$ be open, $f : U \to \mathbb{R}^n$, $a \in U$ and $\rho > 0$ such that $D_\rho(a) \subset U$. Let $i, j \in \{1, \dots, m\}$ and suppose that $D_i D_j f(x)$ and $D_j D_i f(x)$ exist for all $x \in D_\rho(a)$ and are continuous at a. Then

$$D_i D_j f(a) = D_j D_i f(a)\,.$$

Proof. Wlog, may assume $n = 1$. If $i = j$, then trivial. Let $i \neq j$, and define

$$g_{ij}(t) = f(a + te_i + te_j) - f(a + te_i) - f(a + te_j) + f(a)\,.$$

Then for each fixed t, define $\phi : [0, 1] \to \mathbb{R}$ by

$$\phi(s) = f(a + ste_i + te_j) - f(a + ste_i)\,.$$

Then we get

$$g_{ij}(t) = \phi(1) - \phi(0)\,.$$

By the mean value theorem and the chain rule, there is some $\theta \in (0, 1)$ such that

$$g_{ij}(t) = \phi'(\theta) = t(D_i f(a + \theta te_i + te_j) - D_i f(a + \theta te_i))\,.$$

Now apply the mean value theorem to the function

$$s \mapsto D_i f(a + \theta te_i + ste_j)\,,$$

there is some $\eta \in (0, 1)$ such that

$$g_{ij}(t) = t^2 D_j D_i f(a + \theta te_i + \eta te_j)\,.$$

We can do the same for g_{ji} and find some $\tilde{\theta}, \tilde{\eta} \in (0, 1)$ such that

$$g_{ji}(t) = t^2 D_i D_j f(a + \tilde{\theta} te_i + \tilde{\eta} te_j)\,.$$

Since $g_{ij} = g_{ji}$, we get

$$t^2 D_j D_i f(a + \theta te_i + \eta te_j) = t^2 D_i D_j f(a + \tilde{\theta} te_i + \tilde{\eta} te_j)\,.$$

Divide by t^2 and take the limit $t \to 0$. By the continuity of the partial derivatives, we get

$$D_j D_i f(a) = D_i D_j f(a)\,.$$

□

Hence, whenever the second derivatives are continuous, the order does not matter. We can alternatively state this result as follows.

Corollary. If $f : U \to \mathbb{R}^n$ is differentiable in U such that $D_i D_j f(x)$ exists in a neighbourhood of $a \in U$ and are continuous at a. Then Df is differentiable at a and

$$D^2 f(a)(u, v) = \sum_j \sum_i D_i D_j f(a) u_i v_j$$

is a symmetric bilinear form.

Proof. This follows from the fact that continuity of second partials implies differentiability, and the symmetry of mixed partials. □

Finally, we conclude with a version of Taylor's theorem for multivariable functions.

Theorem 7.33 (Second-order Taylor's theorem). Let $f : U \to \mathbb{R}$ be C^2, i.e. $D_i D_j f(x)$ exist and are continuous for all $x \in U$. Let $a \in U$ and $D_r(a) \subset U$. Then

$$f(a+h) = f(a) + Df(a)h + \frac{1}{2}D^2 f(a)(h,h) + E(h)\,,$$

where $E(h) = o(\|h\|^2)$.

Proof. Consider the function

$$g(t) = f(a+th)\,.$$

Then the assumptions tell us g is twice differentiable. By the 1D Taylor's theorem, we know

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(s)$$

for some $s \in [0,1]$. In other words,

$$\begin{aligned} f(a+h) &= f(a) + Df(a)h + \frac{1}{2}D^2 f(a+sh)(h,h) \\ &= f(a) + Df(a)h + \frac{1}{2}D^2 f(a)(h,h) + E(h)\,, \end{aligned}$$

where

$$E(h) = \frac{1}{2}(D^2 f(a+sh)(h,h) - D^2 f(a)(h,h))\,.$$

Then

$$\|E(h)\| \le \frac{1}{2}\left\|D^2 f(a+sh) - D^2 f(a)\right\|\|h\|^2\,.$$

By the continuity of the second derivative, we get

$$\left\|D^2 f(a+sh) - D^2 f(a)\right\| \to 0$$

as $h \to 0$. Hence, $E(h) = o(\|h\|^2)$. □

Appendices: IA Analysis

A The Real Number System

Definition A.1. A set F together with two binary operations, say addition and multiplication, is a *field* if elements in F satisfy the axioms:

1. *Closure.* $\forall a, b \in F$, $a + b$, $ab \in F$;
2. *Associativity.* $\forall a, b, c \in F$, $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$;
3. *Commutativity.* $\forall a, b \in F$, $a + b = b + a$, $ab = ba$;
4. *Distributivity.* $\forall a, b, c \in F$, $a(b + c) = ab + ac$;
5. *Additive identity.* $\exists e \in F$ such that $\forall a \in F$, $e + a = a$;
6. *Multiplicative identity.* $\exists u \in F$, $u \neq e$ such that $\forall a \in F$, $ua = a$;
7. *Additive inverse.* $\forall a \in F$, $\exists b \in F$ such that $a + b = e$;
8. *Multiplicative inverse.* $\forall a \in \mathbb{F}$, $a \neq e$, $\exists b \in \mathbb{F}$ such that $ab = u$.

Definition A.2. A *totally ordered set* is a set X with a relation $<$ that satisfies

1. *Transitivity.* if $x, y, z \in X$ and $x < y$, $y < z$, then $x < z$;
2. *Trichotomy.* if $x, y \in X$, exactly one of $x < y$, $x = y$ and $y < x$ holds.

Definition A.3. An *ordered field* is a field $\mathbb{F}$ with a relation $<$ that makes $\mathbb{F}$ into an ordered set such that

1. if $x, y, z \in \mathbb{F}$ and $x < y$, then $x + z < y + z$;
2. if $x, y, z \in \mathbb{F}$, $x < y$ and $z > 0$, then $xz < yz$.

Definition A.4. Let X be an ordered set and let $A \subseteq X$. An *upper bound* for A is an element $x \in X$ such that $\forall a \in A$, $a \leq x$. If A has an upper bound, then A is *bounded above*.

An upper bound x for A is the *least upper bound* or *supremum* if $\forall y < x$, $\exists a \in A$ such that $a > y$. This is denoted as

$$\sup A = x\,.$$

The *greatest lower bound* or *infimum*, denoted $\inf A$ is defined similarly.

Example. Let $X = \mathbb{Q}$. The set $\{x | x^2 < 2\}$ is bounded above, but has no supremum as $\sqrt{2} \notin \mathbb{Q}$.

Definition A.5. An ordered set X has the *least upper bound property* if every non-empty subset of X that is bounded above has a supremum.

Definition A.6. The *real numbers* are an ordered field with the least upper bound property.

B Convergence of Sequences and Series

B.1 Sequence

Definition B.1. A *sequence* is a function $\mathbb{N} \to \mathbb{R}$ (or $\mathbb{C}$).

Definition B.2. If a_n is a sequence of complex numbers, we say $a_n \to z$ as $n \to \infty$ if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$, $|a_n - z| < \epsilon$, written

$$\lim_{n\to\infty} a_n = z\,.$$

We say that the sequence *converges* to z. Otherwise, we say a_n *diverges*.

Lemma B.3 (Archimedean property). Let $\mathbb{F}$ be an ordered field with the least upper bound property, then

(i) the set $\{1, 2, 3, \dots\}$ is not bounded above;

(ii) the sequence $a_n = 1/n \to 0$ as $n \to \infty$.

Definition B.4. A sequence a_n is *bounded* if $\exists R \in \mathbb{R}$ such that $\forall n \in \mathbb{N}$,

$$|a_n| \le R\,.$$

A sequence is *eventually bounded* if $\exists R \in \mathbb{R}$ and $N \in \mathbb{N}$ such that $\forall n \ge N$,

$$|a_n| \le R\,.$$

Lemma B.5. An eventually bounded sequence is bounded.

Lemma B.6. Let a_n and b_n be sequences with limits a and b,

(i)
$$\lim_{n\to\infty} (a_n + b_n) = a + b\,;$$

(ii)
$$\lim_{n\to\infty} (a_n b_n) = ab\,;$$

(iii) if $b \neq 0$, then
$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}\,.$$

Theorem B.7 (Sandwich rule). Let a_n and b_n be sequences that both converge to x. Suppose that $a_n \le c_n \le b_n$ for every n. Then $c_n \to x$.

B.2 Monotone Sequences Property

Definition B.8. A sequence a_n is *increasing* if $a_n \le a_{n+1}$ for all n, and is *strictly increasing* if $a_n < a_{n+1}$.

(Strictly) decreasing sequences are defined analogously. A (strictly) increasing or decreasing sequence is called *(strictly) monotone*.

Definition B.9. An ordered field has the *monotone sequences property* if every increasing sequence that is bounded above converges.

Lemma B.10. Least upper bound property $\iff$ monotone sequences property.

Lemma B.11 (Nested intervals property). Let $\mathbb{F}$ be an ordered field with the monotone sequences property. Let $I_1 \supseteq I_2 \supseteq \dots$ be closed bounded non-empty intervals. Then

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset\,.$$

Definition B.12. Let a_n be a sequence. A *subsequence* of a_n is a sequence of the form $a_{n_1}, a_{n_2}, \ldots$, where $n_1 < n_2 < \ldots$.

Theorem B.13 (Bolzano–Weierstrass theorem). Let $\mathbb{F}$ be an ordered field with the monotone sequences property (in particular $\mathbb{F} = \mathbb{R}$). Then every bounded sequence has a convergent subsequence.

Proof. Let u_0 and v_0 be a lower and upper bound respectively for a sequence a_n. By repeated bisection, we can find a sequence of intervals $[u_0, v_0] \supseteq [u_1, v_1] \supseteq [u_2, v_2] \supseteq \ldots$ such that $v_n - u_n = (v_0 - u_0)/2^n$, and such that each $[u_n, v_n]$ contains infinitely many terms of a_n.

By the nested intervals property, $\bigcap_{n=1}^{\infty} [u_n, v_n] \neq \emptyset$. Let x belong to the intersection. Pick a subsequence $a_{n_1}, a_{n_2}, \ldots$ such that $a_{n_k} \in [u_k, v_k]$. We will show that $a_{n_k} \to x$.

Let $\epsilon > 0$. By the Archimedean property, we can find K such that $v_K - u_K = (v_0 - u_0)/2^K \leq \epsilon$. This implies that $[u_K, v_K] \subseteq (x - \epsilon, x + \epsilon)$, since $x \in [u_K, v_K]$. Then $\forall k \geq K$, $a_{n_k} \in [u_k, v_k] \subseteq [u_K, v_K] \subseteq (x - \epsilon, x + \epsilon)$. So $|a_{n_k} - x| < \epsilon$.

B.3 Cauchy Sequences

Definition B.14. A sequence a_n is *Cauchy* if for all $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that whenever $p, q \geq N$, we have $|a_p - a_q| < \epsilon$.

Example. The sequence $\{3, 3.1, 3.14, 3.141, \ldots\}$ in $\mathbb{Q}$ is Cauchy as the terms become arbitrarily close, but it is not convergent as its limit $\pi \notin \mathbb{Q}$.

Lemma B.15. Every convergent sequence is Cauchy.

Theorem B.16 (General principle of convergence). Let $\mathbb{F}$ be an ordered field with the monotone sequences property. Then every Cauchy sequence of $\mathbb{F}$ converges.

Definition B.17. An ordered field in which every Cauchy sequence converges is called *complete*.

Hence $\mathbb{R}$ is a complete ordered field.

Lemma B.18. An ordered field that is complete and has the Archimedean property satisfies monotone sequences property.

B.4 Limit Supremum and Infimum

Definition B.19. Let a_n be a bounded sequence. Its *limit supremum* is

$$\limsup_{n\to\infty} a_n \coloneqq \lim_{n\to\infty} \left(\sup_{m \geq n} a_m \right),$$

and its *limit infimum* is

$$\liminf_{n\to\infty} a_n \coloneqq \lim_{n\to\infty} \left(\inf_{m \geq n} a_m \right).$$

C Convergence of Series

Definition C.1. Let a_n be a real sequence. For each N, define the *N-th partial sum*

$$S_N \coloneqq \sum_{n=1}^{N} a_n \,.$$

If the sequence S_N converges to some limit s, then we say that the *series* converges to s, written

$$\sum_{n=1}^{\infty} a_n = s\,.$$

Lemma C.2. If $\sum_{n=1}^{\infty} a_n$ converges, then $a_n \to 0$.

The converse is not true.

Example (Harmonic series). Let $a_n = 1/n$. Then

$$S_{2^n} - S_{2^{n-1}} = \frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^n} \geq \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

and so $S_{2^n} \geq S_1 + n/2$. The series is therefore unbounded despite $a_n \to 0$.

Example (Geometric series). Let $|\rho| < 1$, then we have

$$\sum_{n=0}^{N} \rho^n = \frac{1-\rho^{N+1}}{1-\rho}$$

and so

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho}\,.$$

Lemma C.3 (Comparison test). Let a_n and b_n be non-negative sequences, and suppose that $\exists C \in \mathbb{R}$ and $N \in \mathbb{N}$ such that $\forall n \geq N$, $a_n \leq C b_n$. Then if $\sum b_n$ converges, then so does $\sum a_n$.

C.1 Absolute Convergence

Definition C.4. A series $\sum a_n$ *converges absolutely* if the series $\sum |a_n|$ converges.

Lemma C.5. If $\sum a_n$ converges absolutely, then $\sum a_n$ converges.

Definition C.6. A series $\sum a_n$ *converges unconditionally* if the series $\sum_{n=1}^{\infty} a_{\pi(n)}$ converges for any bijection $\pi : \mathbb{N} \to \mathbb{N}$, i.e. no matter how we re-order the elements of a_n, the sum still converges.

Theorem C.7. Absolute convergence $\iff$ unconditional convergence. Moreover, if $\sum a_n$ converges absolutely, then for any bijection $\pi : \mathbb{N} \to \mathbb{N}$,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\pi(n)}\,.$$

C.2 Convergence Tests

Lemma C.8 (Alternating series test). Let a_n be a decreasing sequence of non-negative real numbers, and $a_n \to 0$, then

$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n$$

converges.

Lemma C.9 (Ratio test). We have two versions:

(i) If $\exists c < 1$ and $\exists N$ such that $\forall n \geq N$,

$$\frac{|a_{n+1}|}{a_n} \leq c\,,$$

then $\sum a_n$ converges.

(ii) If $\exists \rho \in (-1, 1)$ such that

$$\frac{a_{n+1}}{a_n} \to \rho\,,$$

then $\sum a_n$ converges.

Lemma C.10 (Condensation test). Let a_n be a decreasing non-negative sequence. Then $\sum a_n$ converges if and only if

$$\sum_{k=1}^{\infty} 2^k a_{2^k}$$

converges.

We also have the integral test although we have not yet defined integrals.

Lemma C.11 (Integral test). Let $f : [1, \infty) \to \mathbb{R}$ be a decreasing non-negative function. Then $\sum_{n=1}^{\infty} f(n)$ converges if and only if

$$\int_1^{\infty} f(x)\,\mathrm{d}x < \infty\,.$$

C.3 Complex Series

Most definitions so far carry over unchanged to the complex numbers. Two exceptions are least upper bound and monotone sequences, because the complex numbers do not have an ordering.

A Cauchy sequence in $\mathbb{C}$ is convergent, as its real part and imaginary part converge respectively. The Bolzano–Weierstrass theorem still holds. The nested-intervals property has a nested-box property as a complex analogue. The ratio test also works in $\mathbb{C}$.

Lemma C.12 (Abel's test). Let $a_1 \geq a_2 \geq \cdots \geq 0$, and suppose that $a_n \to 0$. Let $z \in \mathbb{C}$ such that $|z| = 1$ and $z \neq 1$. Then

$$\sum_{n=1}^{\infty} a_n z^n$$

converges.

D Continuous Functions

D.1 Continuous Functions

Definition D.1. Let $\Omega \subseteq \mathbb{R}$ and $f : \Omega \to \mathbb{R}$. Then f is *continuous* at a point $a \in \Omega$ if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\forall y \in \Omega$, $|y - a| < \delta$, we have $|f(y) - f(a)| < \epsilon$.

f is *continuous* if it is continuous at every $a \in \Omega$.

Lemma D.2. Let $\Omega \subseteq \mathbb{R}$ and $f, g : \Omega \to \mathbb{R}$ be continuous, then

(i) $f + g$ is continuous;

(ii) fg is continuous;

(iii) if $g \neq 0$, f/g is continuous.

Lemma D.3. Let $\Omega_1, \Omega_2 \subseteq \mathbb{R}$ and $f : \Omega_1 \to \Omega_2$, $g : \Omega_2 \to \mathbb{R}$. Then if f and g are continuous, $g \circ f : \Omega_1 \to \mathbb{R}$ is continuous.

Theorem D.4 (Maximum value theorem). Let $[a, b]$ be a closed interval in $\mathbb{R}$ and let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is bounded and attains its bounds, i.e. $\exists x, y \in [a, b]$ such that $f(x) = \sup_{[a,b]} f$ and $f(y) = \inf_{[a,b]} f$.

Theorem D.5 (Intermediate value theorem). Let $a < b \in \mathbb{R}$ and let $f : [a, b] \to \mathbb{R}$ be continuous. Suppose that $f(a) < c < f(b)$. Then there exists an $x \in (a, b)$ such that $f(x) = c$.

D.2 Continuous Induction

Proposition D.6 (Continuous induction v1). Let $a < b$ and let $\Omega \subseteq [a, b]$ have the following properties:

(i) $a \in \Omega$;

(ii) if $x \in \Omega$ and $x \neq b$, then $\exists y \in \Omega$ with $y > x$;

(iii) if $\forall \epsilon > 0$, $\exists y \in \Omega$ such that $y \in (x - \epsilon, x]$, then $x \in \Omega$.

Then $b \in \Omega$.

It can also be formulated as below.

Proposition D.7 (Continuous induction v2). Let $\Omega \subseteq [a, b]$ and suppose

(i) $a \in \Omega$;

(ii) if $[a, x] \subseteq \Omega$ and $x \neq b$, then there exists $y > x$ such that $[a, y] \subseteq \Omega$;

(iii) If $[a, x) \subseteq \Omega$, then $[a, x] \subseteq \Omega$.

Then $\Omega = [a, b]$.

E Differentiability

E.1 Limits

Definition E.1. Let $\Omega \subseteq \mathbb{R}$ and let $f : \Omega \to \mathbb{R}$. Let $a \in \Omega$. We say the *limit* of f at $x = a$ is l, written

$$\lim_{x \to a} f(x) = l$$

if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\forall 0 < |x - a| < \delta$, we have $|f(x) - l| < \epsilon$.

Proposition E.2. Let $f(x) \to l$ and $g(x) \to m$ as $x \to a$. Then as $x \to a$,

(i) $f(x) + g(x) \to l + m$;

(ii) $f(x)g(x) \to lm$;

(iii) $f(x)/g(x) \to l/m$ if $m \neq 0$.

Definition E.3. A function $f : \Omega \to \mathbb{R}$ is *differentiable* at $a \in \Omega$ with a *derivative* λ if

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lambda$$

or equivalently

$$\lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = \lambda \,.$$

This is denoted

$$f'(a) = \lambda \text{ or } \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=a} = \lambda \,.$$

From this, we can recursively define the multiple derivatives of f.

Proposition E.4. f is differentiable $\iff$ $f(x + h) = f(x) + hf'(x) + o(h)$.

Lemma E.5 (Sum and product rule). Let f, g be differentiable at x, then $f + g$ and fg are differentiable at x, with

$$(f + g)'(x) = f'(x) + g'(x)$$
$$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$$

Lemma E.6 (Chain rule). If f is differentiable at x and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x with derivative $g'(f(x))f'(x)$.

Lemma E.7 (Quotient rule). If f and g are differentiable at x and $g(x) \neq 0$, then f/g is differentiable at x with derivative

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2} \,.$$

Lemma E.8. If f is differentiable at x, then it is continuous at x.

Theorem E.9. Let $f : [a, b] \to [c, d]$ be differentiable on (a, b), continuous on $[a, b]$ and strictly increasing. Suppose that $f'(x)$ never vanishes and $f(a) = c$ and $f(b) = d$. Then f has an inverse and for each $y \in (c, d)$, g is differentiable at y with derivative $1/f'(g(y))$.

E.2 Differentiation Theorems

Theorem E.10 (Rolle's theorem). Let f be continuous on $[a, b]$ and differentiable on (a, b). Suppose that $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.

Corollary (Mean value theorem). Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that

$$f'(x) = \frac{f(b) - f(a)}{b - a} \,.$$

Theorem E.11 (Local version of inverse function theorem). Let f be a function with continuous derivative on (a, b). Let $x \in (a, b)$ and suppose that $f'(x) \neq 0$. Then there is an open interval (u, v) containing x on which f is invertible. Let $g = f^{-1}$, then

$$g'(f(z)) = \frac{1}{f'(z)}$$

for every $z \in (u, v)$.

This says that if f has a non-zero derivative, then it has an inverse locally and the derivative of the inverse is $1/f'$.

Theorem E.12 (High-order Rolle's theorem). Let f be continuous on $[a, b]$ and n-times differentiable on an open interval containing $[a, b]$. Suppose that

$$f(a) = f'(a) = \cdots = f^{(n-1)}(a) = f(b) = 0\,,$$

then $\exists x \in (a, b)$ such that $f^{(n)}(x) = 0$.

Proof. Induct on n. The $n = 1$ case is the Rolle's theorem.

Suppose now we have $k < n$ and $x_k \in (a, b)$ such that $f^{(k)}(x_k) = 0$. Since $f^{(k)}(a) = 0$, we can find $x_{k+1} \in (a, x_k)$ such that $f^{(k+1)}(x_{k+1}) = 0$ by Rolle's theorem. The result follows by induction.□

Corollary. Suppose f and g are both differentiable on an open interval containing $[a, b]$ and that $f^{(k)}(a) = g^{(k)}(a)$ for $k = 0, 1, \ldots, n-1$, and also $f(b) = g(b)$. Then there exists $x \in (a, b)$ such that $f^{(n)}(x) = g^{(n)}(x)$.

Proof. Apply generalised Rolle's theorem to $f - g$. □

For any f, we can find a polynomial p of degree at most n that satisfies the conditions for g, i.e. a p such that $p^{(k)}(a) = f^{(k)}(a)$ for $k = 0, 1, \ldots, n-1$ and $p(b) = f(b)$. From this, we have the Taylor's theorem.

Theorem E.13 (Taylor's theorem with the Lagrange form of remainder). Let f be continuous on $[a, a+h]$ and n-times differentiable on $(a, a+h)$. Then

$$f(a+h) = f(a) + hf'(a) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{h^n}{n!} f^{(n)}(x)$$

for some $x \in (a, a+h)$. The last term is the error term in Lagrange form.

E.3 Complex Differentiation

Definition E.14. Let $f : \mathbb{C} \to \mathbb{C}$. Then f is differentiable at $z \in \mathbb{C}$ with a derivative $f'(z)$ if

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists and equals $f'(z)$. Equivalently,

$$f(z+h) = f(z) + hf'(z) + o(h)\,.$$

F Complex Power Series

Definition F.1. A *complex power series* is a series of the form

$$\sum_{n=0}^{\infty} a_n z^n$$

where $z \in \mathbb{C}$ and $a_n \in \mathbb{C}$ for all n. It is a function of z when it converges.

Lemma F.2. Suppose $\sum a_n z^n$ converges and $|w| < |z|$, then $\sum a_n w^n$ converges absolutely.

From this, we can define the radius of convergence of a complex power series.

Definition F.3. The *radius of convergence* of a power series $\sum a_n z^n$ is

$$R \coloneqq \sup\left\{ |z| \;\middle|\; \sum a_n z^n \text{ converges} \right\}.$$

Lemma F.4. The radius of convergence of $\sum a_n z^n$ is

$$R = \frac{1}{\limsup(|a_n|)^{1/n}}\,.$$

F.1 Exponential and Trigonometric Functions

Definition F.5. The *exponential function* is

$$\mathrm{e}^z \coloneqq \sum_{n=0}^{\infty} \frac{z^n}{n!}\,,$$

which converges on all of $\mathbb{C}$.

Definition F.6. Having two sequences a_n and b_n, their convolution is the sequence c_n defined by

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0\,.$$

The relevance of this is

$$\sum_{n=0}^{N} c_n z^n = \left(\sum_{n=0}^{N} a_n z^n\right)\left(\sum_{n=0}^{N} b_n z^n\right).$$

Theorem F.7. Let $\sum a_n$ and $\sum b_n$ be two absolutely convergent series and let c_n be the convolution of sequences a_n and b_n. Then the series $\sum c_n$ converges absolutely and

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right).$$

Definition F.8. We define *sine* and *cosine* to be

$$\sin z \coloneqq \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots$$
$$\cos z \coloneqq \frac{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}}{2} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} + \dots$$

Proposition F.9.

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{e}^z = \mathrm{e}^z$$
$$\frac{\mathrm{d}}{\mathrm{d}z}\sin z = \cos z$$
$$\frac{\mathrm{d}}{\mathrm{d}z}\cos z = -\sin z$$

Definition F.10. Let x be the smallest positive real number such that $\cos x = 0$. $\pi \coloneqq 2x$.

From this, we have our familiar induction formulae, sum angle rule etc. for trigonometric functions.

Theorem F.11. Let $\sum a_n z^n$ be a power series with radius of convergence R. For $|z| < R$, let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ and } g(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}\,.$$

Then f is differentiable with a derivative g.

F.2 Hyperbolic Trigonometric Functions

Definition F.12. Define hyperbolic sine and cosine to be

$$\cosh z = \frac{\mathrm{e}^z + \mathrm{e}^{-z}}{2} = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots$$
$$\sinh z = \frac{\mathrm{e}^z - \mathrm{e}^{-z}}{2} = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots$$

Proposition F.13.

$$\frac{\mathrm{d}}{\mathrm{d}z}\cosh z = \sinh z$$
$$\frac{\mathrm{d}}{\mathrm{d}z}\sinh z = \cosh z$$

G The Riemann Integral

G.1 Riemann Integral

Definition G.1. Let $[a, b]$ be a closed interval. A *dissection* $\mathcal{D}$ of $[a, b]$ is a sequence $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. The *mesh* of a dissection $\mathcal{D}$ is $\sup(x_{i+1} - x_i)$.

Definition G.2. Given a dissection $\mathcal{D}$ of a interval $[a, b]$ and $f : [a, b] \to \mathbb{R}$. The *lower sum* and *upper sum* are defined by

$$U_{\mathcal{D}}(f) \coloneqq \sum_{i=1}^{n}(x_i - x_{i-1}) \sup_{x\in[x_{i-1},x_i]} f(x)\,,$$
$$L_{\mathcal{D}}(f) \coloneqq \sum_{i=1}^{n}(x_i - x_{i-1}) \inf_{x\in[x_{i-1},x_i]} f(x)\,.$$

We often use the shorthand

$$M_i = \sup_{x\in[x_{i-1},x_i]} f(x)\,,\ m_i = \inf_{x\in[x_{i-1},x_i]} f(x)\,.$$

Lemma G.3. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two dissections of $[a, b]$. Then

$$U_{\mathcal{D}_1} f \geq L_{\mathcal{D}_2} f\,.$$

Definition G.4. The *upper integral* is

$$\overline{\int_a^b} f(x)\,\mathrm{d}x = \inf_{\mathcal{D}} U_{\mathcal{D}} f\,.$$

The *lower integral* is

$$\underline{\int_a^b} f(x)\,\mathrm{d}x = \sup_{\mathcal{D}} L_{\mathcal{D}} f\,.$$

If these are equal, then we call their common value the *Riemann integral* of f, denoted

$$\int_a^b f(x)\,\mathrm{d}x\ .$$

A function is *Riemann integrable* if its Riemann integral exists.

Proposition G.5 (Riemann integrability criterion). Let $f : [a, b] \to \mathbb{R}$. Then f is Riemann integrable if and only if for any $\epsilon > 0$, there exists a dissection $\mathcal{D}$ such that

$$U_{\mathcal{D}} - L_{\mathcal{D}} < \epsilon\,.$$

Lemma G.6 (Linearity of integration). Let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable and $\lambda, \mu \in \mathbb{R}$. Then $\lambda f + \mu g$ is Riemann integrable and

$$\int_a^b \lambda f + \mu g\,\mathrm{d}x = \lambda \int_a^b f\,\mathrm{d}x + \mu \int_a^b g\,\mathrm{d}x\ .$$

Proposition G.7 (Additivity). Let $f : [a, c] \to \mathbb{R}$ be integrable with $b \in (a, c)$, then

$$\int_a^c f \, \mathrm{d}x = \int_a^b f \, \mathrm{d}x + \int_b^c f \, \mathrm{d}x \ .$$

Proposition G.8. If $f, g : [a, b] \to \mathbb{R}$ are integrable, then fg is integrable.

Theorem G.9. A continuous function f on a closed bounded interval $[a, b]$ is Riemann integrable.

Theorem G.10. Let $f : [a, b] \to \mathbb{R}$ be monotone, then f is Riemann integrable.

Proof. Let $\epsilon > 0$. Let $\mathcal{D}$ be a dissection of mesh less than $\frac{\epsilon}{f(b)-f(a)}$. Then

$$\begin{aligned} U_\mathcal{D} f - L_\mathcal{D} f &= \sum_{i=1}^n (x_i - x_{i-1})(f(x_i) - f(x_{i-1})) \\ &\leq \frac{\epsilon}{f(b) - f(a)} \sum_{i=1}^n f(x_i) - f(x_{i-1}) \\ &= \epsilon \ . \end{aligned}$$

□

Lemma G.11. A function $f : [a, b] \to \mathbb{R}$ that is bounded and continuous on (a, b) is integrable.

An example is $\int_0^1 \sin \frac{1}{x} \, \mathrm{d}x$. It is well-behaved apart from $x = 0$, so it is integrable.

Corollary. Every piecewise continuous and bounded function on $[a, b]$ is integrable.

Theorem G.12 (The fundamental theorem of calculus).

(i) Let $f : [a, b] \to \mathbb{R}$ be continuous and for $x \in [a, b]$, define

$$F(x) = \int_a^x f(t) \, \mathrm{d}t \ .$$

Then F is differentiable and $F'(x) = f(x)$.

(ii) Let $f : [a, b] \to \mathbb{R}$ be a differentiable function, and suppose f' is integrable, then

$$\int_a^b f'(t) \, \mathrm{d}t = f(b) - f(a) \, .$$

Proof.

(i)

$$\frac{F(x + h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t) \, \mathrm{d}t \ .$$

Let $\epsilon > 0$. Since f is continuous at x, then there exists δ such that $|y - x| < \delta$ implies $|f(y) - f(x)| < \epsilon$. If $|h| < \delta$, then

$$\begin{aligned} \left| \frac{1}{h} \int_x^{x+h} f(t) - f(x) \right| &= \left| \frac{1}{h} \int_x^{x+h} (f(t) - f(x)) \, \mathrm{d}t \right| \\ &\leq \frac{1}{|h|} \left| \int_x^{x+h} |f(t) - f(x)| \, \mathrm{d}t \right| \\ &\leq \frac{\epsilon |h|}{|h|} \\ &= \epsilon \ . \end{aligned}$$

(ii) Let $\mathcal{D}$ be a dissection $x_0 < \cdots < x_n$. write

$$f(b) - f(a) = \sum_{i=1}^{n} f(x_i) - f(x_{i-1}) \,.$$

For each i, there exists $u_i \in (x_{i-1}, x_i)$ such that $f(x_i) - f(x_{i-1}) = (x_i - x_{i-1})f'(u_i)$ by the mean value theorem. So

$$f(b) - f(a) = \sum_{i=1}^{n} (x_i - x_{i-1}) f'(u_i) \,.$$

We know that $f'(u_i)$ is between $\sup\limits_{x \in [x_{i-1}, x_i]} f'(x)$ and $\inf\limits_{x \in [x_{i-1}, x_i]} f'(x)$, so

$$L_{\mathcal{D}} f' \le f(b) - f(a) \le U_{\mathcal{D}} f' \,.$$

As f' is integrable and $\mathcal{D}$ is arbitrary, the upper and lower sums can get arbitrarily close to the integral, so

$$f(b) - f(a) = \int_a^b f'(t) \,\mathrm{d}t \ .$$

□

Theorem G.13 (Integration by parts). Let $f, g : [a, b] \to \mathbb{R}$ be differential with f', g' integrable. Then

$$\int_a^b f(x) g'(x) \,\mathrm{d}x = [f(x)g(x)]_a^b - \int_a^b f'(x) g(x) \,\mathrm{d}x \ .$$

Theorem G.14 (Taylor's theorem with the integral form of remainder). Let f be $(n+1)$ times differentiable on $[a, b]$ with $f^{(n+1)}$ continuous. Then

$$f(b) = f(a) + (b-a)f'(a) + \cdots + \frac{(b-a)^n}{n!} f^{(n)}(a) + \int_a^b \frac{(b-t)^n}{n!} f^{(n+1)}(t) \,\mathrm{d}t \ .$$

G.2 Improper integral

Definition G.15. Suppose there is a function $f : [a, b] \to \mathbb{R}$ such that $\forall \epsilon > 0$, f is integrable on $[a + \epsilon, b]$. We define the *improper integral*

$$\int_a^b f(x) \,\mathrm{d}x = \lim_{\epsilon \to 0} \int_{a+\epsilon}^b f(x) \,\mathrm{d}x$$

if the limit exists, even if the integral does not exist.

Similarly, we may define

$$\int_a^\infty f(x) \,\mathrm{d}x = \lim_{b \to \infty} \int_a^b f(x) \,\mathrm{d}x$$

if it exists.